University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 218

Subjective completion beliefs and the demand for post-secondary education

Johannes S. Kunz and Kevin E. Staub

February 2016

Subjective completion beliefs and the demand for post-secondary education*

Johannes S. Kunz
University of Zurich

Kevin E. Staub
University of Melbourne, and IZA

February 22, 2016

Abstract

The outcome of pursuing a post-secondary educational degree is uncertain. A student might not complete a chosen degree for a number of reasons, such as academic insufficiency or financial constraints. Thus, when considering whether to invest in post-secondary education, students must factor in their completion probability into their decision. We study the role of this uncertainty in educational choices using students' subjective beliefs about completing a post-secondary education, which were elicited prior to students' completing secondary education. We relate these subjective completion probabilities to their subsequent educational choices and outcomes using representative survey data from Germany. Following the students over time, we find that the initial beliefs are predictive of intentions to invest in education, actual subsequent educational investments, and degree completion. We assess the heterogeneity of the impact across different educational paths. After controlling for academic ability, we find that subjective beliefs are most relevant in choosing a vocational education. In addition to reduced form models, we estimate a structural choice model of sequential investment in education that allows for unobserved tastes and preferences for education and forward-looking behavior. The results confirm the influence of subjective completion beliefs on choosing a post-secondary education.

Keywords: Subjective beliefs, Educational completion uncertainty, Human Capital Investment.
JEL classification: I21, I26, J24

*Addresses for correspondence Johannes S. Kunz: University of Zürich, Department of Economics, Zürichbergstrasse 14, CH-8039 Zürich; johannes.kunz@econ.uzh.ch. Kevin E. Staub: Department of Economics, 111 Barry Street, The University of Melbourne, 3010 VIC, Australia; kevin.staub@unimelb.edu.au.
We thank Joseph Altonji, Orazio Attanasio, Uschi Backes-Gellner, Dan Black, Timo Boppart, Gregory Crawford, Thomas Dohmen, Ernst Fehr, Joe Hirschberg, Jamie Gloor, Uli Kaiser, Ed Lazear, Edwin Leuven, Jenny Lye, Philippe Ruh, Florian Schaffner, Carmit Segal, Andreas Steinhauer, Steven Stillman, Sabrina Studer, Rainer Winkelmann, Joachim Winter, and seminar participants at Kloster Kappel, Luzern, Melbourne, Schaffhausen, Zurich, the Ski and Labor Seminar, Laax, and the Workshop of Personnel Economics and Economics of Education, Lech, for helpful comments and suggestions, and Daniel Auer and Luca Tonizzo for very helpful research assistance.

1 Introduction

Educational choice is one of the most important career decisions young adults have to make, and one that must be made under partial information. Students not only face the difficulty of having to predict labor market prospects for several educational degrees —an endeavoring task even for highly distinguished scholars (Manski, 1993)— but also the challenge of having to predict their own returns to each of these degrees. When choosing an educational track, students further need to foresee their own abilities for that new educational environment and their chances of succeeding in such an environment. Understanding the role of uncertainty in individuals' post-secondary educational choices is essential for designing effective educational policies. For instance, if students' expectations are misaligned, providing additional information can be a cost-effective measure to enhance educational choices, and eventual career success.

In this study, we assess the role of uncertainty about completing an educational degree in young adults' education choices. We show how subjective beliefs about completing a post-secondary education, elicited while in secondary education are important predictors of post-secondary educational aspirations, enrollment, and completion. We find that both beliefs and educational aspirations are strongly related to academic ability and personality traits. Conversely, actual enrollment and completion depend to a larger extent on family characteristics, the state of the local labor market, and the regional supply and demand in the post-secondary education market, which suggests a potential for informational policy interventions. In addition, we assess effect heterogeneity across post-secondary educational tracks. Academic ability appears to be the main driver of subjective beliefs in choosing a university education. In contrast, subjective beliefs of those choosing a vocational education seem to be driven by other characteristics. Finally, we develop and estimate a structural educational choice model that accounts for unobserved preferences for post-secondary education, forward-looking behavior, and the sequentiality of choices. We find that the subjective completion beliefs are most decisive for adolescents with low academic ability and weak preferences for education; a group that has largely been ignored in the present literature.

The context of this study is the secondary and post-secondary educational system in Germany, which is renowned for its well-functioning apprenticeship system. Apprenticeship systems are now tested and implemented in several countries (including the US, cf. President Barack Obama's State of the Union Address, 2014) motivated by the low youth unemployment rates observed in countries with apprenticeship systems. In Germany, when finishing secondary education at the age of 16 to 17, young adults choose between dropping out of education, investing in an apprenticeship, or continuing with general education that enables them to enroll in an university. In addition to analyzing the demand for

education under conditions of uncertain outcomes, we study how the demand varies across educational tracks, which is relevant to policy makers who aim to introduce apprenticeship systems.

In general, when studying choice under uncertainty, researchers have to assume how expectations are formed (Manski, 2004). Most commonly, researchers impose rational expectations; e.g., that individuals' predictions, usually about future wage distributions, are unbiased. In the context of individuals' educational choices it is important to note that even if students were able to accurately predict the future wage distribution, their (perceived) internal rates of return —the rates upon which they act— might be very different from the aggregate returns.[1] By far, the most widely used alternative is to use direct measures of elicited subjective beliefs, which circumvents these problems (Manski, 2004). Several studies show that the use of elicited expectation data can be superior to those constructed using rational expectation assumptions, and they are meaningful measures in educational choice models (e.g., Attanasio and Kaufmann, 2014; Huntington-Klein, 2015*b*; Stinebrickner and Stinebrickner, 2012; Zafar, 2011*a*). Although the literature on educational decision making under uncertainty using elicited subjective beliefs is rapidly growing, we advance the literature in important dimensions. We assess the role of prior subjective beliefs formed in secondary education in a representative population survey and follow these adolescents over time until they complete their post-secondary education.[2] Second, much of the existing literature on the demand for post-secondary education focuses on investment, rather than on aspirations or completion.[3] Thus, we assess each of these three outcomes while also accounting explicitly for uncertainty in students' choices.

In contrast to the existing literature that investigates the uncertainty about wages or the likelihood of unemployment, our main focus is on completion uncertainty. Although some theoretical work includes completion uncertainty (e.g., Altonji, 1993; Comay, Melnik and Pollatschek, 1973; Manski, 1989), there is little empirical work in this area. Theoretical studies emphasize the sequentiality of the educational decisions and that "[d]ifferences in dropout probabilities may be more important than differences in ex post payoffs in determining the ex ante return to attending a particular school,"

[1]Several approaches were proposed to circumvent rational expectations. Early approaches based on structural assumptions that distinguish *ex ante* from *ex post* returns include Carneiro et al. (2003), Cunha, Heckman and Navarro (2005), and Cunha and Heckman (2007). Their framework is also applied recently in Foley, Gallipoli and Green (2014). Another approach is to include measures of uncertainty within the expected wage functions of Roy-type selection models; for example, Mazza (2014) introduces the (rational expectation) variance of earnings and Fossen and Glocker (2014) include risk preferences.

[2]Also related to our study is the evolving literature on college major choice using subjective beliefs: Arcidiacono, Hotz and Kang (2013), Arcidiacono et al. (2014), Hastings et al. (2015), Huntington-Klein (2015*c*), Stinebrickner and Stinebrickner (2014*a*), and Wiswall and Zafar (2015*a*). In Germany there are no majors, as students specialize at the beginning of their studies. However, we follow a similar approach as these studies by allowing for selection into different educational tracks.

[3]One reason is that the data on completion is necessarily incomplete: individuals can always come back and acquire more education. For a detailed discussion of educational completion, see Turner (2004) and Bound and Turner (2011). Notable exceptions are Venti and Wise (1983) and Light and Strayer (2000). Similarly, the literature on aspirations is still comparatively small, although it has been growing recently (e.g., Christofides et al., 2015; Wiswall and Zafar, 2015*b*; Zachary and Zafar, 2015).

(Altonji, 1993, p74).[4] This hypothesis is empirically supported by Hussey and Swinton (2011), based on a predicted likelihood of completion. However, such predicted completion probabilities are limited in that they are only a crude proxy for the subjective beliefs on which people act. We contribute to this literature by integrating elicited subjective completion probabilities into a sequential model of educational choice. In this respect, our analysis is most closely related to Wiswall and Zafar (2015*a*), which also uses students' subjective completion beliefs. Our research addresses complementary questions such as how the choice process differs for adolescents not enrolled in college and how these beliefs relate to actual completion. Our paper is the first to study subjective completion beliefs assessed before the end of secondary education in a population survey in the context of a detailed educational investment model.

One way in which completion uncertainty affects educational choice is by simply amplifying *ex ante* wage uncertainty. However, completion uncertainty may have important consequences beyond that general channel. For example, various non-pecuniary aspects have been shown to be relevant to educational choice (see Oreopoulos and Salvanes, 2011, for a recent summary). In order to benefit from them, staying in the chosen educational path and/or completing the degree might be crucial. For instance, studies using elicited subjective beliefs about labor market prospects consistently find the (non-financial) consumption value of education or major-specific unobserved tastes to be the main drivers of educational choices (i.e., Huntington-Klein, 2015*a*; Wiswall and Zafar, 2015*a*).[5] Such preference-related factors are not affected by pure labor market uncertainty, but they can be affected by completion uncertainty.[6] Our results also point to unobserved preferences for a post-secondary education that play a substantial role in students' choices.

Our study is also closely related to the literature on learning about one's own academic ability (or preferences).[7] The central finding in this literature is that learning about one's own ability is based mainly on academic ability conveyed by students' grade point averages [GPA] (e.g., Stinebrickner and Stinebrickner, 2012, 2014*b*; Zafar, 2011*b*). Although these studies offer a valuable

[4] Manski (1989) raises an important point by clarifying that drop-out rates are not necessarily undesirable from a social planner's point of view: since educational outcomes are uncertain, schooling should be evaluated based on *ex ante* returns rather than on *ex post* success rates.

[5] Similar evidence comes from more structural approaches that do not rely on subjective beliefs. For instance, D'Haultfoeuille and Maurel (2013) use a sophisticated Roy model and find non-pecuniary aspects to be predominant in educational choice.

[6] Evidence whether the provision of information about the labor market induces students to invest more in education is mixed, which can be interpreted as broadly in line with our view that there is more to uncertainty than pure wage uncertainty. Supporting evidence comes from developing countries, for instance, see Jensen (2010) for evidence from Dominican Republic and Nguyen (2008) for Madagascar. Oreopoulos and Dunn (2013) find that high school students in Canada update their beliefs in the context of an information experiment. Yet in Finland, Kerr et al. (2014) find that —while students do update their beliefs— there is no significant effect on enrollment; similar results are reported in Fryer (2013). Assessing students' choice process in more detail is therefore highly valuable.

[7] Bulman (2015) shows that providing young adults with better information about their own ability impacts enrollment and college graduation. He finds that important factors other than aptitude deter college attendance, which might be explained by subjective beliefs about educational outcomes.

assessment of the subjective beliefs at various points in time and in great detail, thus far they have focused on single institutions rather than a representative sample. Milla (2014) adds to and supports the generalizability of the previous findings by studying aspiration updating in responses to changes in GPA using a population survey of college students. Still, such a design imposes a sample selection. By exclusively focussing on college students it ignores young adults who dropped out of education because they were less optimistic about their educational prospects. We contribute to this literature by assessing initial subjective beliefs prior to college enrollment in a representative survey population and by providing evidence on both beliefs and educational aspirations at this early stage. Our evidence supports and extends Zafar's presumption that "prior belief[s] [at the start of college] continue[s] to be important. In attempting to understand the choice of college majors, it might be useful to focus on students at earlier stages of their schooling (for example, in high school) and analyze their subjective beliefs" (Zafar, 2011*b*, p339f).[8]

Throughout our analyses we account for personality skills, which have been highlighted as main determinants of educational success (see Almlund et al., 2011; Borghans et al., 2006, and references therein). In particular, we show how subjective beliefs relate to the Big Five personality measures, risk attitudes, and locus of control, all of which are now ubiquitous in economic applications (see, for example, Borghans et al., 2006; Caliendo, Cobb-Clark and Uhlendorff, 2015; Dohmen et al., 2010). Of special interest to our design is the locus of control, as Coleman and DeLeire (2003) hypothesize that students with a more internal locus of control (i.e., students who believe their actions affect their outcomes) have higher subjective beliefs about their own returns to education, which increases their efforts and investments in their human capital. Our results support the hypothesis that one's locus of control affects educational choices via subjective beliefs.

Finally, our study is related to recent contributions assessing the role of subjective beliefs as a mediator and a potential explanation of educational differentials in parental unemployment (Pinger, 2015), family background (Keller and Neidhöfer, 2014), or gender and migration (Tolsma, Need and De Jong, 2010). Our framework might prove useful in studying the mediating role of subjective beliefs, since it integrates investment in both secondary and tertiary education jointly in both reduced-form and structural models.

In sum, the main contribution of this study is to provide a better understanding of uncertainty in educational choices and a broad assessment of subjective completion beliefs of young adults. Our analyses include how beliefs are determined and how beliefs relate to intentions to invest in education,

[8]Due to data limitations, we do not examine subjective beliefs at multiple time points. A detailed analysis of the process behind learning about one's own ability and the evolution of subjective beliefs is beyond the scope of this study, but remain key questions for future research.

actual investments, and degree completion. We explicitly account for the sequentiality of choices and forward-looking behavior of individuals. Moreover, we relate students' beliefs to individual characteristics, family background, personality skills, regional labor and education market conditions, and unobserved tastes and preferences for education. The remainder of this study proceeds as follows: In Section 2, we describe the institutional features of the educational system in Germany and present the data we use. In Section 3, we assess determinants of subjective completion beliefs. In Section 4, we relate the beliefs to educational outcomes and present how the impact of the subjective beliefs varies with selection on observables and unobservables. In Section 5, we presents effect heterogeneity across different educational tracks, and in Section 6 we develop and estimate a structural model of sequential educational choice. Section 7 concludes our paper, briefly summarizing our key findings.

2 Institutional setting, data, and descriptive statistics

Institutional setting

A simplified version of Germany's educational system is depicted in Figure 1, in which we briefly summarize the system's key features that are relevant to our analysis (more information can be found in Wölfel and Heineck, 2012).

— — — Figure 1 about here — — —

The German educational system is characterized by early tracking, which takes place after grade 4 (elementary school), at age 9 to 11 years.[9] Based on grades and teachers' recommendations, the children are tracked into three streams according to their academic ability.[10] The statistical agency in Germany (Statistisches Bundesamt, 2014, p27) reported that in 2012, 10% of children were assigned to the lower track, 19% to the intermediate track, 40% to the upper track (high school), and the remaining children visited other, so-called comprehensive schools that essentially follow the same structure without separating the children.

At the time of entering the survey population, the young adults —ages 16 to 17 years— are in the midst of deciding upon a professional education according to their track. Students completing lower or intermediary tracks have the opportunity to apply for and to start a profession-specific apprenticeship or a vocational education.[11] Although investing directly in an apprenticeship is the dominant path, the young adults can alternatively enroll in a consecutive school-track that leads to the *university entrance*

[9] With the exceptions of Berlin and Brandenburg, which track after grade 6 (ages 11 to 13 years).

[10] The binding nature of these recommendations varies across states.

[11] Students who started an apprenticeship before entering the survey population are excluded from our analyses. However, in 2011, only 10.6% started an apprenticeship before the age of 17 years (Statistisches Bundesamt, 2013, p17).

qualification (German: *Abitur*), an equivalent of a high school degree.[12] This high school degree can also be a valuable asset for students who do not want to attend an university. When applying for highly competitive apprenticeship positions, students with a high school degree typically have better chances compared to their peers who completed a lower track. Some apprenticeship positions are even exclusively available to such students. In 2010, 20.9% of the newly signed apprenticeship contracts went to students holding a high school degree (Statistisches Bundesamt, 2011, p1004). Thus, we model this path separately and refer to it as tertiary apprenticeship.

The decision to start an apprenticeship is somewhat different for students already enrolled in high school. In principle, they can also drop out to start an apprenticeship or continue their high school education and after finishing go on to university or a tertiary apprenticeship.[13] Yet, their default choice is certainly different as they are already enrolled in high school and they do not have to make an active choice to enroll.[14] In sum, 4% of the class of 2011 dropped out without a degree, 17% completed the lower track, 36% the intermediary track, and 43% obtained a high school degree (Statistisches Bundesamt, 2013, p7). As a final remark, it is important to realize that in Germany, an apprenticeship degree has a high standing and a reputation similar to a university degree —especially when acquired after completing high school.

Summing up, in the subsequent analysis we distinguish between the four most commonly taken education paths in Germany, which we index by j. The student can choose to drop out ($j = 0$), invest in an apprenticeship directly after completing either the lower or intermediate school track ($j = 1$), or continue schooling in high school. After completing high school, the student can decide whether to invest in an apprenticeship ($j = 2$) or continue to university studies ($j = 3$).

Data sources

Our primary data source is the German Socio-Economic Panel [SOEP]. We focus on young adults, ages 16 to 17 years, who have newly entered the survey population by answering the youth questionnaire between 2000 and 2013. The SOEP is a household panel that provides a rich set of parental background information. We use all available waves of data collection to follow the young adults over time up to 14 years. Additionally, we combine the individual-level data with regional labor market information and educational supply and demand measures based on 96 geographic regions, which we will

[12] Due to the limited time horizon of our sample we focus on early investment. The possibility of visiting complementary courses that allow students to go to university after apprenticeship completion is not modeled separately. We discuss the implications of this for the interpretation of our results below.

[13] Additionally, students could also drop out after completing high school, but this rarely occurs in practice (see also Fossen and Glocker, 2014). Note that here university subsumes universities of applied sciences. While it would be interesting to consider those separately, we have to leave this to future research due to our current sample size.

[14] We, therefore, include the indicator variable "In high school with 17" in all regressions. We also estimated the regressions of our main Table 3 separately —fully saturated in this variable— for the two groups and present the results in Appendix Table A6.

refer to as Ror (for their German name *Raumordnungsregionen*).[15] All regional information is matched according to the individual's residency when answering the youth questionnaire, and lagged by one year to avoid endogeneity or reverse causality. Unless stated otherwise, we only use variables assessed in the youth questionnaire to avoid any biases from conditioning on outcomes (Angrist and Pischke, 2009, p64f).[16]

Sample selection

As stated above, we exclude all individuals who have already started an apprenticeship. Moreover, we exclude students with missing information in the core variables: subjective belief, GPA, and educational status. All other missing information are included along with corresponding indicator variables for missing observations. This selection results in a sample size of 3,610 individual observations. In the longitudinal analysis, we additionally require at least 2 years of information to assess the end of secondary education and the start of a post-secondary education (reducing the observations to 2,116), and to assess educational completion, we restrict the sample to students who responded for at least 5 years of data collection (1,372).[17]

Descriptive statistics

Our main variable of interest is the subjective completion belief, p_i, that was assessed by the following question:

> Think about your future in your job and private life: how probable is it, in your opinion, that the following events will occur?
>
> [Please check off a probability on the scale from 0% to 100%.]
>
> *You successfully finish your vocational training or university studies?*[18]

There are two caveats about how the question is assessed. First, the question is only elicited once. Second, the question does not elicit beliefs for every possible counterfactual education. We discuss the implications of these issues later in the estimation results. For now, we focus on the role of initial beliefs in the combined effect of any post-secondary education. In this way, the question directly relates to the outcomes that we assess. In Figure 2, we plot histograms of the subjective beliefs by students' intentions to invest in education. Intention to invest is a self-reported measure of educational aspiration

[15] A map of the Ror's is provided in Appendix A, Figure A1. The data source is INKAR 2012 provided by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR, 2013). For more information, see Pinger (2015) who also uses this additional data source. Moreover, we add the number of universities (higher learning institutions) as a proxy for distance to university provided by the statistical agency of Germany (Statistisches Bundesamt).

[16] We develop a structural model below in order to account for sequential decision-taking and to avoid associated biases.

[17] More information on missing values and the construction of the variables can be found in the Appendix, Table A7.

[18] Students could answer on an eleven point scale. The exact wording in German is:
Wenn Sie sich einmal Ihre berufliche und private Zukunft vorstellen: Wie wahrscheinlich ist es, dass die folgenden Entwicklungen eintreten werden? [Stufen Sie bitte jeweils die Wahrscheinlichkeit auf einer Skala ein, die von 0 Prozent bis 100 Prozent geht.] *Ihre Ausbildung oder Ihr Studium erfolgreich abschließen?*

that asks students to indicate which further educational degree (if any) they plan to complete. It is constructed analogously to our outcome variable: 0 refers to no further educational aspiration; 1, to apprenticeship; 2, to tertiary apprenticeship (high school and apprenticeship); and 3, to university studies.[19]

— — — Figure 2 about here — — —

Overall, the German students appear to be confident about finishing a post-secondary education, as most of the adolescents report a probability above 50%. The distributions of students implicitly aspiring to a high school degree (implied either by tertiary apprenticeship or university, Panels C and D) are very similar in shape. Yet, the subjective beliefs of young adults who aspire to a university degree are more concentrated and slightly shifted to the right compared to their high school counterparts who aim for a tertiary apprenticeship position. The mode of the distribution of students who intend to start an apprenticeship without finishing high school lies at 100%. Finally, students with no educational aspiration display a much larger spread in their beliefs. In what follows, it is important to keep in mind that most variation —and the bulk of the students beliefs— are located between 60 and 100%.

Analogously, Table 1 presents descriptive statistics by aspiration level for our baseline sample.[20]

— — — Table 1 about here — — —

At the bottom of the table, we present the sample shares of the intentions to invest: Most students want to complete an apprenticeship, followed by university studies. A substantial share wants to complete a high school education and an apprenticeship (tertiary apprenticeship), and roughly 10% do not aspire to any professional education. It is reassuring that the sample statistics are broadly consistent with the population statistics presented above (Statistisches Bundesamt, 2013).

The individuals who aspire to a university education are on average the most confident about successfully completing their post-secondary education, and have the lowest standard deviation. However, all young adults who have any educational aspirations exhibit a similar level of completion beliefs —which are close to 80%— as opposed to those without educational aspirations. The fact that all students have positive beliefs, even the ones who do not plan to invest in further training or education, can be rationalized in a simple expected utility framework where students weight their utility from

[19]That means that, for consistency with our outcome variables, students who want to enroll in an apprenticeship first and then continue with supplementary courses that prepare for university are subsumed into the apprenticeship category. Moreover, we also cannot distinguish in detail between students who first want to complete an apprenticeship and then a high school degree, without aiming to go university. However, this path is neither optimal from a human capital investment perspective nor one that is commonly taken in Germany.

[20]More information on the construction of the variables can be found in the table notes. Unconditional descriptive statistics for the various subsamples considered in the analysis below are presented in the Appendix, Table A1.

education by their beliefs about their completion probabilities and report their aspirations based on their highest expected utility. This is also the interpretation we pursue in the following analysis.

Some interesting patterns emerge when relating educational aspirations to our three measures of academic ability: Aspirations are increasing in the grade point average [GPA].[21] Prior track recommendations at the age of 10 years seem to be a good indicator for the aspirations up to 7 years later, which could either be caused by a well-working ability streaming or a manifestation of students' expectations as a result of early-tracking. Interestingly, having no educational aspirations occurs in all tracks, and the largest share of students without aspirations is found in high school. This could be explained by a default effect, as the survey elicits these aspirations at a time when students not enrolled in high school have to make an active decision as opposed to their high school counterparts who can follow their track and decide after obtaining a high school degree.

We assess the adolescents' personality by locus of control, risk attitudes, and the Big Five personality inventory.[22] Educational aspirations are positively associated with the locus of control, which measures to what extent a person believes her life is under her own control. Among the standard Big Five inventory, aspirations increase with openness, agreeableness, and extraversion but they are less monotonically related to conscientiousness or neuroticism. Unconditionally, aspiring to a university degree is positively associated with risk attitudes.

Individuals' characteristics and family backgrounds are captured by their gender, number of siblings, whether they are second-generation immigrants (persons whose parents were both born in a foreign country), whether at least one parent has a college education, is currently unemployed, and the logarithm of the net household income. Aspirations tend to be higher among males, children from smaller families, natives, persons with employed and college-educated parents or with a higher household income.

The regional labor and education market (Ror) characteristics relevant for the students' choices set are a mix of (exogenous) educational supply and demand shifters. We use the cyclical component of the youth unemployment rate, and the number of apprenticeship positions, students, high school graduates, and universities in the region. Throughout, the aspirations are increasing with the local labor and education market characteristics as expected; only children with no educational aspirations

[21] The GPA refers to the student's average of the German and Math grade, which is standardized over the sample population we present in our main results from Table 3. We also standardize the GPA within school track in Appendix Table A4 to show that the choice of standardization does not drive the results. We further standardize GPA within federal states to show that different grading levels do not affect the results, see Appendix Table A5.

[22] We standardize all the principal components of the personality variables (all but risk attitudes, which are assessed by one question only), small deviations from (0,1) result from the missing values which do not enter the standardization but are set to 0 afterwards. The locus of control has been developed by Rotter (1966), the Big Five inventory by Costa and McCrae (1992) and validated in the SOEP version by Hahn, Gottschling and Spinath (2012). Risk attitudes have been introduced and extensively studied by Dohmen et al. (2011) and references therein.

tend to have no clear ordering. In the following analysis, we will also account for region and year of first questioning (which is roughly identical to students' age).[23]

3 Determinants of subjective completion beliefs

To analyze how the variables we discussed in the previous section relate to subjective completion beliefs, we estimate OLS regressions of the model

$$p_i = x_i'\beta^p + \upsilon_i, \tag{1}$$

where i indexes individuals, p_i is the subjective completion belief, x_i are varying sets of explanatory variables with corresponding vector of coefficients β^p, and υ_i is an unobserved error term.

The estimates are presented in Table 2.[24] In Column (1), the beliefs are explained solely by academic ability. In Column (2), we add the personality measures; in Column (3), individual and family and individual background characteristics; and, finally, in Column (4), regional measures, year and region fixed effects.

— — — Table 2 about here — — —

The explained variation, as measured by the adjusted R^2, increases substantially only when academic ability and personality measures are included, but stays relatively unaffected when adding individual and family characteristics, or fixed effects and regional characteristics. The joint significance tests for subsets of variables reported at the bottom panel of the table give analogous results: Academic ability and personality characteristics are highly significant across all regressions; individual and family characteristics are jointly significant; labor market characteristics, regional and time effects are not. Since the labor market coefficient estimates are neither jointly nor individually significant, we omitted these estimates from the table.

Looking at the determinants individually, all academic ability variables are consistently positive and significant. Somewhat surprisingly, already being enrolled in high school does not alter students' subjective completion beliefs. This might be due to this effect being conditional on track recommendation. As hypothesized by Coleman and DeLeire (2003), the locus of control is a very important

[23] For some of the regressions, the number of students within a state is too small. To obtain consistent samples, we use a broader grouping by dividing Germany into the following 5 regions (and an indicator for missing values). Southern Germany: Baden-Wuerttemberg, Bavaria; Eastern Germany: Berlin, Brandenburg, Saxony, Saxony-Anhalt, Mecklenburg-Western Pomerania; Central Germany: Hesse, Thuringia; Western Germany: North Rhine-Westphalia, Rhineland-Palatinate, Saarland; Northern Germany: Bremen, Hamburg, Lower Saxony, Schleswig-Holstein. We present analogous results of our main Table 3 in Appendix Table A4 were we use federal states fixed effects, as the jurisdiction over educational policies are on the federal state level. The results are qualitatively the same.

[24] Note that our dependent variable is a fraction. In the Appendix, Table A2, we present fractional response regressions (as in Papke and Wooldridge, 1996, 2008). The results are virtually indistinguishable from the OLS estimates.

determinant of subjective completion beliefs throughout the regressions, both in magnitude and significance.[25] Risk attitudes do not matter once family characteristics are accounted for. Our regressions indicate that among the Big Five measures of personality, conscientiousness is the most influential in shaping subjective beliefs. This finding highlights the importance of conscientiousness for educational outcomes, as is consistently found in the literature (see, *inter alia*, Borghans, Meijers and Ter Weel, 2006). While we find little evidence that openness or neuroticism influence completion beliefs, extraversion has a coefficient which is about half as large as conscientiousness, and the effect of agreeableness is about half as large as extraversion.

On average, females seem to have lower subjective completion beliefs. This estimate is, however, only marginally significant (at least conditional on personality and academic ability). Household income is positively and significantly related to subjective completion beliefs. Being a second-generation immigrant is significantly negatively associated with subjective beliefs. However, the significance vanishes after including regional determinants. This suggests a segregation effect, with immigrants being located in less economically and educationally active areas. The other covariates are insignificant and mostly very small in magnitude.

4 Subjective completion beliefs and educational outcomes

In this section, we turn to our central question of how subjective completion beliefs measured at age 17 years relate to intended investments in education, actual investments in education, and, finally, educational degree attainment. To fix ideas, let the individual i's utility u_{ij} from choosing an uncertain post-secondary educational track ($j \geq 1$) be

$$u_{ij} = \begin{cases} \mu_{ij} + \varepsilon_{ij} & \text{with probability } p_{ij} \\ \bar{\mu}_{ij} + \varepsilon_{ij} & \text{with probability } (1 - p_{ij}) \end{cases}, \tag{2}$$

where p_{ij} is the subjective completion belief, μ_{ij} ($\bar{\mu}_{ij}$) is the utility from (not) completing, and ε_{ij} is an utility component unaffected by completion. The associated expected utility is

$$\begin{aligned} U_{ij} &= p_{ij}\mu_{ij} + (1 - p_{ij})\bar{\mu}_{ij} + \varepsilon_{ij} \\ &= \bar{\mu}_{ij} + p_{ij}(\mu_{ij} - \bar{\mu}_{ij}) + \varepsilon_{ij}. \end{aligned} \tag{3}$$

[25] Caliendo, Cobb-Clark and Uhlendorff (2015) also find a strong link between subjective beliefs and the locus of control in the realm of job search among the unemployed.

Hence, adolescents get a baseline utility from attending a particular educational track $\bar{\mu}_{ij}$. The subjective completion belief p_{ij} weights the utility differential between completing and not completing an educational track either up or down. Since not investing in an educational track does not involve educational uncertainty, its utility is simply

$$U_{i0} \quad = \quad \mu_{i0} + \varepsilon_{i0}, \text{ with certainty.} \tag{4}$$

In this section, we assess the investment in any post-secondary education $U_{ij} = U_i$ for $j \geq 1$, against not investing U_{i0}. The subjective belief p_i therefore corresponds directly to the question in the survey. A student prefers to invest in education if $U_i > U_{i0}$; where, by standard normalization, $\mu_{i0} = 0$. Taking averages across individuals, adding covariates x_i that measure observed preferences and skills, and assuming that $\nu_i = \varepsilon_i - \varepsilon_{i0}$ follows a standard normal distribution, we estimate probit models of the form

$$d_i \quad = \quad 1[\alpha p_i + x_i'\beta^d + \nu_i > 0]. \tag{5}$$

We consider three binary outcomes d_i. First, whether a student intends to invest in any further education, which is measured concurrently with subjective beliefs at age 17 years. Second, whether a student actually invests in any further education; that is, whether the student started an apprenticeship or, tertiary apprenticeship, or enrolled in a university. This event can be a few months or a few years away. Third, whether a student completes an apprenticeship or university degree, an event that is at least a couple of years away.

When d_i stands for the intention to invest, the expectation of (5) gives $P(U_i > 0)$, so that $\alpha = \mu - \bar{\mu}$. A similar interpretation is possible when d_i represents the second outcome. It then corresponds to the revealed preferences of actual investment in post-secondary education. The interpretation is somewhat different when d_i stands for the third outcome, the completion of a degree. In this case, α gives an indication of the student's ability to incorporate information beyond that in x_i into their forecast of $d_i = 1$. Here, we interpret the subjective beliefs in a similar vein as Finkelstein and McGarry (2006). The adolescents process all their available information in forming their beliefs, meaning that relevant information over and above their subjective beliefs are either not used, not used efficiently, or influence the decision through another channel than subjective completion uncertainty.

— — — Table 3 about here — — —

Table 3 contains the estimation results. Vertical panels (A) to (E) present the probit regressions of

the subjective completion beliefs on the different educational outcomes. In each panel, we report the estimated coefficients, robust standard errors (in parentheses), average marginal effects (in squared brackets), pseudo R_n^2 for the model estimated with and without p_i, and sample statistics for the respective subsamples. Columns (1) to (4) contain the simple probit estimates of the educational outcomes on the subjective beliefs and varying sets of covariates: The specification in Column (1) contains, apart from p_i, only an indicator of whether the student is currently in high school, region and year fixed effects. Thus, in this specification, any other variable acts on the intention to invest in education through its effect on p_i. The next columns progressively control for the sets of academic (Column 2), personality (Column 3), and family and labor market variables (Column 4). We turn to the results in Columns (5)-(8) at the end of this section.

Panel (A) contains results corresponding to the intention to invest in any post-secondary education. Unsurprisingly, uncertainty appears to be important for aspirations: The coefficients on subjective beliefs are large and highly significant throughout the probit regressions. The average marginal effects are economically relevant. In the most parsimonious specification, increasing the subjective beliefs in the population by one standard deviation increases intentions to invest in post-secondary education by 2.7 percentage points (0.14×0.198), which is very large given that only 9.2% of students do not intend to invest in a post-secondary education. This changes little if we condition on increasing sets of background characteristics commonly considered in the literature. A one standard deviation change results in an increase of 2.1 percentage points using all background characteristics. Moreover, the increase in the pseudo R^2 when including the subjective beliefs is similar to the increase when adding both personality and family background. Thus, we find subjective beliefs are strongly related to intended behavior, a result consistent with Huntington-Klein (2015*b*).

Several explanations can account for this observed correlation between beliefs and aspirations; therefore, we examine whether the link from beliefs to intention carries over to revealed preferences in actual investments (at least two years later). Our dependent variable is now an indicator that equals one if the student started any post-secondary education. Panel (B) uses the broadest sample possible for this question. Compared to (A), it only excludes students who are still in school and students who have not completed any subsequent questionnaires two years after the baseline questionnaire at age 17. The average marginal effect is somewhat smaller than for the intentions, ranging from 1.4 to 0.9 percentage points for one standard deviation increase in the subjective beliefs.[26] In Panel (C), we show that similar results are obtained when restricting the sample to students who expressed

[26] Almost all adolescents in Germany start some post-secondary education, 95.6% in our sample, which explains why the average marginal effect for investment is smaller than that for intentions despite an estimated coefficient of similar magnitude.

earlier intentions to invest in post-secondary education. This shows that the subjective completion belief drives not only hypothetical, intended investment, but it also has real behavioral consequences. Compared to the previous results on students' intentions-to-invest, the set of family background and labor market variables explain a larger fraction of the completion belief effect and exhibit a substantial explanatory power.

It is interesting to compare how the subjective completion beliefs relate to actual completion (at least five years later). This can be interpreted as how well the students can predict their future outcomes.[27] The estimation results are given in Panels (D) and (E). The average completion rate is roughly 55%. Unconditionally, a one standard deviation increase in the subjective beliefs increases completion rates in the population by 3.3 percentage points in the overall sample (Panel D), and by 3.6 when conditioning on the sample with positive intentions (Panel E). This decreases to 2.6 and 3.3 percentage points, respectively, when including the full set of individual, family, and regional characteristics. Comparing the coefficients across rows, a notable result is that for degree completion, the set of covariates that affect the coefficient of p_i most is that of the personality measures. Including these variables reduces the estimated coefficient by about 15 to 20 percent. The explanatory power of personality, family background, and labor market characteristics are substantial. This suggests that the students do not optimally account for this information when forming their beliefs. Again, the explanatory power of the beliefs is substantial.

Taken together, the results show that subjective completion beliefs formed during secondary education are predictive over a long time horizon for future post-secondary education. The subjective beliefs are predictive even after accounting for a large set of previously identified, important characteristics. In the appendix, we present further results showing the robustness of these findings across a number of alternative specifications. We show that the results are robust to dichotomizing the subjective beliefs to a dummy variable, thus accounting for potential non-linearity as discussed in Pinger (2015). Further, since academic ability is found to be the main determinant of learning about one's own ability in the literature, we use various reasonable standardizations of GPA that account for potential differences in grading across federal states, or within high school versus no high school. We also use a fifth-order polynomial to show that the beliefs do not pick up non-linearities in academic ability. Additionally, we use federal state dummies instead of the region dummies used in the main

[27] Since completing a program and graduating takes some time, we only consider students which we see at least five years after they have taken the youth questionnaire when they indicated their completion beliefs. This further reduces our available sample. Moreover, it is clear that students who were interviewed in earlier years are more likely to have completed their degrees simply by virtue of being in the sample for a longer period of time. However, this mechanism is captured by the year fixed effects, and is therefore unlikely to bias our results. A second concern is that some of the observations are censored: As of the time we observe them, some students have not yet completed their degree, but they might do so in the future. In this sense, our results should be interpreted as representing the average effect of completion beliefs on completion within a given time frame.

specification (cf. footnote 23). Finally, we present separate estimations for students enrolled in high school when answering the youth questionnaire to account for the different default choices discussed above in a completely flexible way.

A remaining concern might be that the uncertainty is confounded with unobserved heterogeneity. We therefore use a bounding strategy for the coefficients by taking potential selection on unobservable tastes and preferences for education into account when estimating (5). We use the approach developed by Altonji, Elder and Taber (2005*a*,*b*, 2008, hereafter, AET). More specifically, we simultaneously estimate the models given in (1) and (5), imposing the following dependence between the error terms:

$$(\nu_i, \upsilon_i) \sim \Phi_2(0, 0, 1, 1, \rho), \tag{6}$$

where $\Phi_2(\cdot)$ denotes the bivariate normal distribution, and its arguments are the two errors' means, variances, and their correlation. In other words, we estimate probit models for all outcomes d_i with p_i as a normal endogenous explanatory variable [denoted probit eev hereafter].[28,29] The bounding is achieved by setting the correlation coefficient ρ to increasing values until the coefficient of the subjective beliefs α tends to zero. Note that Column (4) in combination with Table 2's Column (4) is equivalent to the probit eev with $\rho = 0$. AET argue that the selection-on-observables is a reasonable (upper) bound on the selection-on-unobservables. Therefore, we also estimate the model replacing

$$\rho = \frac{cov(x_i'\beta^d, x_i'\beta^p)}{var(x_i'\beta^d)} \equiv \hat{\rho}^o$$

as a suggestive upper bound. Columns (5) to (7) contain the probit eev estimates using the full set of covariates and ρ constrained to 0.1, 0.3, and 0.5. Finally, Column (8) constrains ρ to be equal to the selection-on-observables $\hat{\rho}^o$. Up to a correlation of 0.3 all coefficients are positive, and for aspirations and intentions they remain statistically significant. This is a sizeable correlation when comparing it to the applications considered in AET. When using the AET bound of selection-on-observables in the last column, the coefficients are all statistically significant and similar in magnitude to those using all covariates and a correlation between 0 and 0.1. This indicates that the results are robust to a sizeable

[28]The corresponding log-likelihood is given by

$$\ln L(d_i, p_i; x_i, \alpha, \beta^d, \beta^p, \rho) = \sum_{i=1}^{n} \ln \Phi \left[(2d_i - 1) \left(\frac{x_i'\beta^d + \alpha p_i + \rho(p_i - x_i'\beta^p)}{\sqrt{1-\rho^2}} \right) \right] + \ln \phi \left(p_i - x_i'\beta^p \right). \tag{7}$$

For more information, see the discussion in Greene (2012, p747f).

[29]In contrast to AET, our main variable is a fraction rather than an indicator. Instead of estimating a bivariate probit, we therefore estimate a probit eev. The use of a continuous normal variable is motivated by the estimation of (1), where we found that it made little difference whether it was estimated by OLS or a fractional response model (cf. Appendix Table A2). As a robustness check we dichotomize the subjective beliefs at $p \geq 70\%$ and estimate bivariate probit regressions as in AET. Estimates for such an approach can be found in the Appendix Table A3. The results are similar but somewhat more conservative, possibly due to the reduced variation.

selection-on-unobservables.

The results presented in this section indicate that the uncertainty of 17-year-old students about completing an educational degree is an important determinant of educational choices and outcomes. While we focused our discussion on the average effect, another effect which is of interest is the one corresponding to the marginal student (a student with an outcome probability of 50%). An increase in the subjective belief by one standard deviation (0.2) for this student would: increase her intention to invest by 5.7 percentage points (i.e., $\Phi(0.2 \times 0.716) - \Phi(0)$), her investment by 6.9 percentage points (5.8 if she stated an intention to invest), and her completion by 2.6 percentage points (3.4 if she stated an intention to invest). In sum, differences in beliefs about being capable of successfully finishing a post-secondary educational degree can explain not only differences in intended future investments in schooling, but in actual investments as represented by enrollment into university or obtaining an apprenticeship position. Moreover, students with higher subjective beliefs are also associated with higher completion rates, even after controlling for several potential confounders and allowing for some selection-on-unobservables.

5 A view at the disaggregated level

To understand how subjective beliefs influence educational choices and outcomes, we proceed with a more disaggregated analysis: different educational choices. In the estimations before, we implicitly assumed the subjective belief measure has the same effect on all educational investments. Clearly, while this is a useful simplification that allows to gauge overall average effects, it might also hide important differences in how completion beliefs explain, say, enrollment in a university program versus enrollment in a vocational training degree. In this section, we separately asses each of the three educational investments: apprenticeship ($j = 1$), tertiary apprenticeship ($j = 2$), and university ($j = 3$). Compared to the previous sections, in which the subjective belief measure corresponded directly to the outcome, for the disaggregated educational tracks one would ideally like to assess the role of counterfactual choices. Unfortunately, these were not asked in the survey. Therefore, we condition on students' aspirations; however, results have to be interpreted more cautiously.

As before, the analysis starts at the level of intended investment. This time we present results from a multinomial probit model with four outcomes. The base category is not having any intention to invest, and the remaining categories are the intention to invest in each of the three educational choices mentioned above. The results are presented in Table 4, whose four columns represent specifications with the increasing sets of covariates discussed previously. We present the χ^2-statistic and corresponding p-value for the likelihood ratio test against a restricted model without subjective beliefs.

— — — Table 4 about here — — —

All coefficients are statistically significant and large in magnitude. The first column shows that the overall average marginal effect of subjective beliefs (in squared brackets) found before stems almost exclusively from the fact that students with higher completion beliefs aspire to a university education. Notice that the sum of the three average marginal effects is roughly equal to the corresponding effect presented in the previous section.[30]

Yet, the second column shows that academic variables, such as GPA, are central in shaping this effect. Once the academic background has been accounted for, the effect of subjective belief works mainly through the apprenticeship channel. The average marginal effect for university is reduced substantially and rendered statistically insignificant. Thus, high GPA has the effect of inducing high completion beliefs, which in turn pushes students towards desiring a university degree. But within a given GPA level, a higher completion belief is positively associated with starting an apprenticeship. Compared to this big shift, the changes resulting from adding personality, family background, and labor market variables are modest (at least beyond its effect through GPA or personality skills). This result is in line with those found in the literature that for college students most of the information is based on measures of academic ability (Milla, 2014; Stinebrickner and Stinebrickner, 2012, 2014*b*; Zafar, 2011*b*). Yet, for those who choose a less theoretical education, other characteristics seem to be more influential.

Turning to the behavioral responses, we estimate separate regressions for the three subsamples according to intended educational choice. That is, we address the question, for example, of how does a higher subjective completion belief increase a student's university enrollment and completion probabilities, given that the student aspired to a university degree. Estimates are reported in Table 5. Columns (1) to (4) display the results for enrollment and (5) to (8) for completion. The three vertical Panels (A) to (C) contain separate probit regressions of each educational track.

— — — Table 5 about here — — —

Panel (A) reports the estimates for investment and completion of an apprenticeship. The results indicate that the effects of subjective beliefs are indeed large and statistically significant, with the average marginal effect ranging from about 10 to 15 percentage points, depending on the set of control variables used. With 883 observations, the subsample with intended investment in apprenticeship is the largest of the three subsamples, accounting for almost half of the total number of observations. The subsample for intended tertiary apprenticeship has only 456 observations. In three out of the

[30] We omit the average marginal effect for the base category: it is equal to minus the sum of the marginal effects of the remaining categories.

four specifications shown in the middle panel, the estimated coefficient is insignificant. The coefficient —and hence the marginal effect— increases as more covariates are controlled for, and only reaches marginal significance in the last column of the panel. The bottom panel, containing the results of the subsample aspiring to a university education, features the opposite pattern. Here, completion beliefs have a large, statistically significant effect on enrollment in university. However, academic background explains almost half of the effect. Adding more sets of control variables further erodes the effect of subjective beliefs on college enrollment.

Comparing the results across the three panels, it appears that the decision to enroll in a post-secondary educational program is related most strongly to subjective completion beliefs for those students aspiring to a university degree (see columns 1 across panels). At the same time, the determinants of these beliefs are mainly related to observable academic and demographic variables for the university-aspiring students. For the two apprenticeship streams, the observable academic and demographics add comparatively little information to the completion beliefs. Moreover, for the tertiary apprenticeship, there even seems to be a negative correlation with these characteristics, but the estimation results are too imprecise to allow for further interpretation.

To conclude this part, we estimate analogous probit regressions for the probability to graduate. The results are depicted in the right-hand-side panel of Table 5. Small sample size issues are a concern, especially for the tertiary apprenticeship graduation regressions. However, the estimates are consistent with our previous results. In particular, the aggregate effect found in the previous section is corroborated in the apprenticeship category. For students who indicated their intentions to invest in an apprenticeship degree at age 17 years, the subjective completion beliefs are highly informative about their actual completion years later. The average marginal effect is close to 30 percentage points —a figure that is reduced to about 20 percentage points after accounting for differences in observables. We cannot estimate precise effects for tertiary apprenticeships. While the point estimates are sizeable, none of them are statistically significant. The results for students aspiring to a university education also echo the previous results. The effect of the completion beliefs at age 17 years is large and statistically significant: a one-standard-deviation change in p_i increases the probability of graduating from university by about 6 percentage points for a student with a baseline graduation probability of 50 percent. Finally (and as before), the available control variables, particularly academic background variables, explain a large portion of this effect.

6 A dynamic model of educational choice

In this section, we conclude our investigation by developing and estimating a model of educational investment along the lines of Taber's (2001) seminal contribution that encompasses three features. First, we allow for the sequential nature of the process: students can only decide whether they want to go to university if they chose to finish high school previously (Altonji, 1993; Comay, Melnik and Pollatschek, 1973). Second, we introduce the dynamics of the optimization process: when deciding whether to go to the labor market or to go to high school, forward-looking students account for the option value of continuing education after finishing high school (Stange, 2012; Trachter, 2015). Finally, we allow for unobserved factors that influence student utilities derived from their choices, which may be correlated across choices and over time, a topic of substantial attention in the returns to education literature (see, e.g., Belzil, 2007; Card, 2001, and references therein).

Model

We consider a stylized two-period model in which students sequentially choose between risky educational paths, as outlined in Figure 3.

— — — Figure 3 about here — — —

Ex ante, students do not know for certain whether they will successfully complete the chosen education track, but they have subjective beliefs, p_i, about finishing. The first period or first stage ($T = 1$) occurs when students finish compulsory education at the age of 17 years. At this point, they face the choice between dropping out of school ($d_{i1} = 0$), investing in an apprenticeship training ($d_{i1} = 1$), or continuing with high school education ($d_{i1} = 2$). A high school degree involves the option value of continuing with tertiary education. Students who choose high school reach the second period ($T = 2$), where they graduate from high school and now have the choice of either investing in a tertiary apprenticeship ($d_{i1} = 2, d_{i2} = 0$) or in a university education ($d_{i1} = 2, d_{i2} = 1$).[31,32]

As mentioned in Section 4, apprenticeships, tertiary apprenticeships, and university all involve uncertainty, which we model according to equations (2)-(4). A key assumption of this approach is that utility can be decomposed into a component that depends on the realization of graduation (μ_{ij} for graduation vs. $\bar{\mu}_{ij}$ else) and an idiosyncratic component unaffected by graduation, ε_{ij}, which captures features such as a preference for attending, say, university irrespective of receiving a degree.

By backward induction, we begin with the students' second stage problem. Students advancing

[31] As noted before, students could also drop out at this point, but this is an extremely rare event in the data and therefore not modeled (see also Fossen and Glocker, 2014).

[32] We only focus on the initial beliefs in shaping young adults' educational choices because we lack a repeated measurement of the subjective beliefs at the end of high school that would allow us to study the learning about ones' own ability in more detail.

to the second stage choose between starting a tertiary apprenticeship ($j = 2$) or going to university ($j = 3$). We denote this choice by d_{i2}, a binary variable where 1 represents choosing university,

$$d_{i2} = \begin{cases} 1 & \text{if } U_{i3} - U_{i2} > 0 \\ 0 & \text{if } U_{i3} - U_{i2} \leq 0 \end{cases}$$

which we specify analogously to equation (4) by

$$U_{i3} - U_{i2} \quad = \quad \alpha_3 p_i + x'_{i,t+1}\beta_3 + \delta_3 \theta_i + \nu_{i3} \equiv z_{i3,t+1} + \nu_{i3}.$$

Here, $\alpha_3 = (\mu_3 - \mu_2) - (\bar{\mu}_3 - \bar{\mu}_2)$, $\nu_{i3} = \varepsilon_{i3} - \varepsilon_{i2}$, and $x_{i,t+1}$ consists of the same covariates considered above, although we include time-varying labor market conditions measured two years after answering the youth questionnaire, which is the time one would need to start a higher education after obtaining a high school degree.[33] This exogenous variation in the decision problem induced by the timing of the events provides an additional source of identification (see, French and Taber, 2011; Taber, 2000, for a discussion on the identification for these models), which has become standard practice in the literature on dynamic models of educational choice (e.g., Heckman et al., 2014; Taber, 2001). To allow for dependence of the unobservables between the two time periods in a flexible way, we add a standard normal random variable θ_i to the utilities, capturing unobserved tastes and preferences for education. We assume that $\nu_{i3} \sim N(0, \sigma_3)$, thus the probability of choosing university is given by

$$P(d_{i2} = 1) \quad = \quad \Phi\left(\frac{z_{i3,t+1}}{\sigma_3}\right),$$

where $\Phi(\cdot)$ represents the univariate normal cdf.

In the first stage, the student has an expectation about her second stage decision (she knows the distribution of ν_{i3}) but does not know her realized value. If students knew their realized ν_{i3} at the time of the first stage, the model would reduce to a simple static polychotomous choice problem, similar to those estimated and reported in Table 4. Thus, the students' expectation about her value of advancing to the second stage, as formed during the first stage, is

$$\mathrm{E}(\max(U_{i3} - U_{i2}, 0)) \quad = \quad \sigma_3 \left[\Phi\left(\frac{z_{i3,t}}{\sigma_3}\right)\frac{z_{i3,t}}{\sigma_3} + \phi\left(\frac{z_{i3,t}}{\sigma_3}\right)\right] \equiv EV_i,$$

and $\phi(\cdot)$ denotes the normal pdf. Now the labor market and educational supply and demand characteristics are measured at time t, one year before the adolescent answers the youth questionnaire

[33] We use students' location at the age of 17 for the region fixed effects, to avoid a bias due to moving.

corresponding to her information set. The difference between high school and drop out utility is then

$$\begin{aligned} U_{iHS} - U_{i0} &= \alpha_2 p_i + x'_{i,t}\beta_2 + \delta_2\theta_i + EV_i + \nu_{iHS} \\ &\equiv z_{iHS,t} + \nu_{iHS}, \end{aligned}$$

which comprises EV_i, the option value of continuing to the second stage. In these types of models, we cannot distinguish between the baseline utility of the second stage tertiary apprenticeship and the utility of high school. The coefficients $\alpha_2, \beta_2, \delta_2$ thus capture the sum of these two effects (while the coefficients in $z_{i3,t}$ correspond to the differences between preferences for university and tertiary apprenticeship).

The apprenticeship utility is

$$U_{i1} - U_{i0} = \alpha_1 p_i + x'_{i,t}\beta_1 + \theta_i + \nu_{i1} \equiv z_{i1,t} + \nu_{i1},$$

where we set $\delta_1 = 1$, a necessary normalization to identify the impact of unobserved heterogeneity, θ_i, on latent utilities. By the bivariate normal assumption on the ν's we can write the probabilities

$$\begin{aligned} P(d_{i1} = 2) &= \Phi_2\left(z_{iHS,t},\ \ z_{iHS,t} - z_{i1,t},\ \ 0.5\right), \\ P(d_{i1} = 1) &= \Phi_2\left(z_{i1,t},\ \ z_{i1,t} - z_{iHS,t},\ \ 0.5\right), \\ P(d_{i1} = 0) &= 1 - P(d_{i1} = 1) - P(d_{i1} = 2). \end{aligned}$$

The individual likelihood contribution, conditional on the unobserved heterogeneity θ_i, is given by

$$\begin{aligned} l_i(\theta_i) &= \{1 - P(d_{i1} = 1) - P(d_{i1} = 2)\}^{\mathbf{1}(d_{i1}=0)} \quad \times \quad \{P(d_{i1} = 1)\}^{\mathbf{1}(d_{i1}=1)} \\ &\times \quad \{P(d_{i1} = 2)[1 - P(d_{i2} = 1)]\}^{\mathbf{1}(d_{i1}=2,d_{i2}=0)} \\ &\times \quad \{P(d_{i1} = 2)P(d_{i2} = 1)\}^{\mathbf{1}(d_{i1}=2,d_{i2}=1)}, \end{aligned} \tag{8}$$

and to obtain the marginal likelihood contribution, we integrate over the distribution of θ,

$$l_i = \int l_i(\theta_i)\phi(\theta_i)d\theta_i,$$

an expression which we approximate by simulation, $\tilde{l}_i$, by taking random draws from the distribution of θ_i. We then maximize the simulated sample log-likelihood $\sum_i \ln \tilde{l}_i$.

Results

The estimation results are depicted in Table 6 in two panels. The left-hand-side panel contains estimates from a constrained version of the model without heterogeneity ($\theta_i = 0$ for all i), whereas the right-hand-side panel contains estimates from the full model with unobserved heterogeneity. Moving from left to right, the columns again contain the expanding set of covariates considered previously. With the exception of the local labor and education market characteristics, all the regressors are time-invariant.[34]

— — — Table 6 about here — — —

Table 6 depicts large and significant estimates for the coefficients of the subjective probabilities in the indices for both $d_{i1} = 1$ and $d_{i1} = 2$. Thus, these results, too, suggest that a higher p_i pushes students away from leaving school without further investments. In particular, the coefficients for $d_{i1} = 2$ suggest that subjective completion beliefs are important determinants of second-stage participation; of completing high school and beginning a tertiary apprenticeship or university studies (which confirms and extends the results found by Pinger, 2015). On the other hand, the coefficients for university are insignificant throughout, and close to zero when accounting for covariates. This indicates that, once in the second stage, the initial subjective beliefs are not informative about the choice of tertiary apprenticeship versus university. A potential explanation for this is belief updating in response to new information revealed by high school grades.

— — — Figure 4 about here — — —

Figure 4 illustrates the role of the option value, EV_i, in shaping the choice probabilities. The figure uses predicted probabilities obtained from the estimated parameters in Column (5) and evaluated at sample means. The left-hand-side panel artificially sets the expected value to zero; that is, we evaluate a constrained model where students ignore the option value of further investment. Thus, we interpret the coefficients from $d_{i1} = 2$ as corresponding only to high school utility, and we assume students neglect the option value of continuing to the second-stage choices. As expected, it can be seen by contrasting the two panels in the figure that the option value decreases the level of the apprenticeship probabilities and increases those of the second-stage choices. But the option value also affects how the probabilities change with the subjective belief, making the gradient on apprenticeship flatter — and even slightly negative for high values of p_i— and the gradient on university steeper. Therefore, the option value in conjunction with the subjective beliefs can play a substantial role in shaping the

[34]The estimate of σ_3 is only identified when time-varying covariates are included (French and Taber, 2011; Taber, 2000). We present it therefore only in Columns (4) and (8).

adolescents' high school investment.[35]

We now turn to the role of the unobserved preferences for post-secondary education or unobserved skills, θ_i. Comparing the two panels of Table 6, we see that all the significant coefficients in the right-hand-side panel, which accounts for such heterogeneity, are somewhat larger than the ones from the left-hand-side. Recall that θ_i has no natural scale, its scale has been fixed such that a unit coefficient in the index corresponds to apprenticeship. The coefficient on the linear index for the baseline second-stage utility is about 0.85 across all specifications (5)-(8) and highly significant. It shows that there is a strong positive correlation between unobserved preferences for apprenticeship and for high school. Unobserved preferences for education are very important for the adolescents investment decisions, as found in the prior literature (e.g. Bulman, 2015; D'Haultfoeuille and Maurel, 2013; Huntington-Klein, 2015*a*; Wiswall and Zafar, 2015*a*). Yet, there is no evidence for differences between unobserved tertiary-apprenticeship-specific skills versus university-specific skills, with the estimated coefficient being virtually zero, potentially a result of preference updating within high school.

Figure 5 further uses the results from Table 6 to visualize how the effect of p_i might differ for different "types" of students.

— — — Figure 5 about here — — —

Based on the estimates of the full specification from Column (8), we define four types by their academic ability level (high versus low GPA) and their unobserved skill level (high versus low θ_i) and plot their predicted choice probabilities against p_i, evaluated at sample means.[36] For students who have high observed and unobserved skills, subjective completion beliefs have negligible effects on investment probabilities. Yet, for adolescents with low unobserved skills (and high GPA), subjective beliefs positively influence all educational tracks. For students with low academic performance, subjective beliefs are more relevant if they have a low preference for education. It is also interesting to note that high GPA (for given level of θ_i) has a much larger effect on investment than high unobserved skills and preferences (for a given level of GPA). This suggests that the subjective beliefs are most relevant for students with low unobserved skills.[37]

In sum, the results from the dynamic sequential model with unobserved heterogeneity shed light on some aspects of educational choice which were masked in the reduced form models of Sections 4 and 5.

[35]The extent of these effects depends on the values of the covariates, which in Figure 4 were set to sample means. In Figure A2 in the Appendix we present graphs where we set all the linear indices $x_i'\beta_j = 0$ and thus obtain effects which are much stronger.

[36]Specifically, we define high and low values of GPA and θ_i as $\Phi^{-1}(0.75)$ and $\Phi^{-1}(0.25)$.

[37]Figure A3 contains a similar graph to Figure 5, but evaluated at $x_i'\beta_j = 0$, thus yielding even more pronounced effects.

One such aspect is that the sequentiality of choices shows that p_i has a highly significant effect on the combined high school and second-stage choice; in contrast, it was difficult to estimate precise effects for p_i in the static reduced form model where all four choices were disaggregated. Furthermore, we have seen that accounting for the option value and for unobserved skills can modify the effect of p_i on the choice probabilities. Speaking more broadly, the structural estimates confirm the main results from the reduced-form estimations presented previously: subjective probabilities contain predictive information for educational investments even after accounting for differences due to an extensive set of controls and unobserved heterogeneity. Additionally, and consistent with the recent literature, throughout the analysis GPA has been shown to be the main driver of subjective beliefs. Thus, learning about one's own ability is largely determined by school grades already before entering a post-secondary education.

7 Conclusion

In this paper, we investigated the role of uncertainty for 17-year-olds on their post-secondary educational outcomes by means of subjective beliefs. Two features of this problem are the young age of the students at the time their subjective completion probabilities were elicited and the long time horizon of the choices to which these measures referred. Both features make this a difficult problem, and it is remarkable that these necessarily crude initial beliefs retain their predictive power over a period of several years. The effects of subjective beliefs on investment intentions and actual investments in any post-secondary education are substantial, remaining so even after controlling for observables. Moreover, subjective beliefs have explanatory power comparable to that of academic and personality variables combined. For the marginal student, a one standard deviation increase in subjective beliefs is associated with a 6 percentage points increase for investment intentions and a 7 percentage points increase for actual enrollment. Finally, the subjective probabilities of completion are also predictive of actual completion, increasing completion probability by 3 percentage points.

When disaggregating the educational tracks and estimating a structural choice model, we find the subjective beliefs most relevant for students who aim for a university degree. This is due to the information revealed by GPA, which broadly confirms results found in the literature. Most notably, we confirm Zafar's (2011*b*) finding that *ex ante* subjective beliefs continue to be important even until the degree is completed. Advancing his findings, we conclude that this is even true for subjective beliefs formed already in or before investing (or staying) in high school. Conditional on the academic ability, the subjective beliefs are most relevant for students who start an apprenticeship, which is largely driven by students with additionally low unobserved skills or preferences for post-secondary education. The literature on subjective beliefs in educational choice has largely ignored these students

and evidence on their learning/decision-making processes are almost non-existent. Our study suggests that these students deserve more attention.

References

Almlund, Mathilde, Angela Lee Duckworth, James Heckman and Tim Kautz. 2011. Personality Psychology and Economics. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 1–181.

Altonji, Joseph G. 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics* 11(1):48–83.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*a*. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources* 40(4):791–821.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*b*. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113(1):151–184.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2008. "Unobservables Assess Bias from Using Selection on Observed Variables to when Evaluating Swan-Ganz." *American Economic Review: Papers and Proceedings* 98(2):345–350.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press, New Jersey.

Arcidiacono, Peter, Joseph V. Hotz, Arnaud Maurel and Teresa Romano. 2014. "Recovering Ex Ante Returns and Preferences for Occupations using Subjective Expectations Data." *National Bureau of Economic Research. Working Paper Series Nr. 20626*.

Arcidiacono, Peter, Joseph V. Hotz and Songman Kang. 2013. "Modeling college major choices using elicited measures of expectations and counterfactuals." *Journal of Econometrics* 166(1):3–16.

Attanasio, Orazio P. and Katja M. Kaufmann. 2014. "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender." *Journal of Development Economics* 109:203–216.

BBSR, Bundesinstitut für Bau-Stadt-und Raumforschung. 2013. "Raumordnungsregionen (ROR)." Accessed 18.12.2015.
`www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Raumordnungsregionen/raumordnungsregionen_node.html`

Belzil, Christian. 2007. "The return to schooling in structural dynamic models: a survey." *European Economic Review* 51(5):1059–1105.

Borghans, Lex, Angela L. Duckworth, James J. Heckman and Bas ter Weel. 2006. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43(4):972–1059.

Borghans, Lex, Huub Meijers and Bas Ter Weel. 2006. "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry* 46(1):2–12.

Bound, John and Sarah E. Turner. 2011. Dropouts and Diplomas. The Divergence in Collegiate Outcomes. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 573–613.

Bulman, Geogre. 2015. "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics* 7(4):1–36.

Caliendo, Marco, Deborah A. Cobb-Clark and Arne Uhlendorff. 2015. "Locus of Control and Job Search Strategies." *Review of Economics and Statistics* 97(1):88–103.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5):1127–1160.

Carneiro, Pedro, Karsten T. Hansen, James J. Heckman and Pedro Carneiro. 2003. "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice." *International Economic Review* 44(2):361–422.

Cebi, Merve. 2007. "Locus of Control and Human Capital Investment Revisited." *Journal of Human Resources* 42(4):919–932.

Christofides, Louis N., Michael Hoy, Joniada Milla and Thanasis Stengos. 2015. "Grades, aspirations and post-secondary education outcomes." *Canadian Journal of Higher Education* 45(1):48–82.

Coleman, Margo and Thomas DeLeire. 2003. "An Economic Model of Locus of Control and the Human Capital Investment Decision." *Journal of Human Resources* 38(3):701–721.

Comay, Yochanan, Arie Melnik and Mosche A. Pollatschek. 1973. "The Option Value of Education and the Optimal Path for Investment in Human Capital." *International Economic Review* 14(2):421–435.

Costa, Paul Jr. and Robert McCrae. 1992. "Four ways five factors are basic." *Personality and Individual Differences* 13(6):653–665.

Cunha, Flavio and James J. Heckman. 2007. "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling." *Labour Economics* 14(6):870–893.

Cunha, Flavio, James J. Heckman and Salvador Navarro. 2005. "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers* 57(2):191–261.

D'Haultfoeuille, Xavier and Arnaud Maurel. 2013. "Inference on an extended Roy model, with an application to schooling decisions in France." *Journal of Econometrics* 174(2):95–106.

Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100(3):1238–1260.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9(3):522–550.

Finkelstein, Amy and Kathleen McGarry. 2006. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review* 96(4):938–958.

Foley, Kelly, Giovanni Gallipoli and David A. Green. 2014. "Ability , Parental Valuation of Education and the High School Dropout Decision." *Journal of Human Resources* 49(4):906–944.

Fossen, Frank M. and Daniela Glocker. 2014. "Stated and revealed heterogeneous risk preferences in educational choice." *IZA DP 7950*.

French, Eric B. and Christopher R. Taber. 2011. "Identification of models of the labor market." *Handbook of Labor Economics* 4:537–617.

Fryer, Roland G. Jr. 2013. "Information and Student Achievement: Evidence from a Cellular Phone Experiment." *NBER Working Papers 19113, National Bureau of Economic Research, Inc.*

Greene, William H. 2012. *Econometric Analysis*. 7 ed. Pearson Education Limited. Edinburgh Gate, Harlow.

Hahn, Elisabeth, Juliana Gottschling and Frank M. Spinath. 2012. "Short measurements of personality - Validity and reliability of the GSOEP Big Five Inventory (BFI-S)." *Journal of Research in Personality* 46(3):355–359.

Hastings, Justine S., Christopher A. Neilson, Anely Ramirez, Seth D. Zimmerman and Seth D. Zimmermann. 2015. "(Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data." *NBER Working Papers 21330, National Bureau of Economic Research, Inc.*.

Heckman, James J., John E. Humphries, Gregory Veramendi and Sergio Urzúa. 2014. "Education, Health and Wages." *NBER Working Papers 19971, National Bureau of Economic Research, Inc.*

Huntington-Klein, Nick. 2015*a*. "Educational Choice: Subjective Data and Consumption Value." *WP*.

Huntington-Klein, Nick. 2015*b*. "Subjective and projected returns to education." *Journal of Economic Behavior & Organization* 117:10–25.

Huntington-Klein, Nick. 2015*c*. ""(Un)informed College and Major Choice": Verification in an Alternate Setting." *Working Paper*.

Hussey, Andrew J. and Omari H. Swinton. 2011. "Estimating the ex ante Expected Returns to College." *American Economic Review: Papers and Proceedings* 101(3):598–602.

Jensen, R. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125(2):515–548.

Keller, Tamás and Guido Neidhöfer. 2014. "Who Dares, Wins? A sibling analysis of tertiary education transition in Germany." *SOEPpapers 713*.

Kerr, Sari P., Tuomas Pekkarinen, Matti Sarvimaki and Roope Uusitalo. 2014. "Educational Choice and Information on Labor Market Prospects: A Randomized Field Experiment." *Working Paper*.

Light, Audrey and Wayne Strayer. 2000. "Determinants of College Completion: School Quality or Student Ability?" *Journal of Human Resources* 35(2):289–308.

Manski, Charles F. 1989. "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon." *Economics of Education Review* 8(4):305–312.

Manski, Charles F. 1993. Adolescent econometricians: How do youth infer the returns to schooling? In *Studies of supply and demand in higher education*, ed. Charles T. Clotfelter and Michael Rothschild. Chicago: University of Chicago Press pp. 43–60.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72(5):1329–1376.

Mazza, Jacopo. 2014. "Does Risk Matter? A Semiparametric Model for Educational Choices in the Presence of Uncertainty." *WP*.

Milla, Joniada. 2014. "Tertiary Education Choices under Uncertainty." *Working Paper*.

Nguyen, Trang. 2008. "Information, role models and precived returns to education." *Mimeo*.

Obama, Barack. 2014. "State of the Union Address." Accessed 18.12.2015.
`http://www.whitehouse.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address`

Oreopoulos, Philip and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1):159–184.

Oreopoulos, Philip and Ryan Dunn. 2013. "Information and College Access: Evidence from a Randomized Field Experiment." *Scandinavian Journal of Economics* 115(1):3–26.

Papke, Leslie E. and Jeffrey M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11(6):619–632.

Papke, Leslie E. and Jeffrey M. Wooldridge. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics* 145(1-2):121–133.

Pinger, Pia R. 2015. "Understanding the Mechanisms behind Intergenerational Effects of Economic Distress." *Working Paper*.

Rotter, Julian B. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs* 80(1):1–28.

Stange, Kevin E. 2012. "An Empirical Investigation of the Option Value of College Enrollment." *American Economic Journal: Applied Economics* 4(1):49–84.

Statistisches Bundesamt. 2011. Die Neue Berufsbildungsstatistik. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2013. Berufsbildung auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2014. Schulen auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Stinebrickner, Ralph and Todd Stinebrickner. 2012. "Learning about academic ability and the college dropout decision." *Journal of Labor Economics* 30(4):707–748.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*a*. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies* 81(1):426–472.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*b*. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics* 32(3):601–644.

Taber, Christopher R. 2000. "Semiparametric identification and heterogeneity in discrete choice dynamic programming models." *Journal of Econometrics* 96(2):201–229.

Taber, Christopher R. 2001. "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies* 68(3):665–691.

Tolsma, Jochem, Ariana Need and Uulkje De Jong. 2010. "Explaining participation differentials in Dutch higher education: The impact of subjective success probabilities on level choice and field choice." *European Sociological Review* 26(2):235–252.

Trachter, Nicholas. 2015. "Stepping stone and option value in a model of postsecondary education." *Quantitative Economics* 6(1):223–256.

Turner, Sarah E. 2004. Going to College and Finishing College Explaining Different Educational Outcomes. In *College Decisions: How Students Actually Make Them and How They Could*, ed. Caroline Hoxby. University of Chicago Press for NBER.

Venti, Steven F. and David A. Wise. 1983. "Individual attributes and self-selection of higher education: College attendance versus college completion." *Journal of Public Economics* 21(1):1–32.

Wiswall, Matthew and Basit Zafar. 2015*a*. "Determinants of college major choice: Identification using an information experiment." *Review of Economic Studies* 82(2):791–824.

Wiswall, Matthew and Basit Zafar. 2015*b*. "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital* 9(2):117–169.

Wölfel, Oliver and Guido Heineck. 2012. "Parental risk attitudes and children's secondary school track choice." *Economics of Education Review* 31(5):727–743.

Zachary, Bleemer and Basit Zafar. 2015. "Intended College Attendance: Evidence from an Experiment on College Returns and Cost." *IZA DP No. 9445*.

Zafar, Basit. 2011*a*. "Can Subjective Expectations Data be used in Choice Models? Evidence n Cognitive Biases." *Journal of Applied Econometrics* 26(3):520–544.

Zafar, Basit. 2011*b*. "How Do College Students Form Expectations?" *Journal of Labor Economics* 29(2):301–348.

Figures & Tables

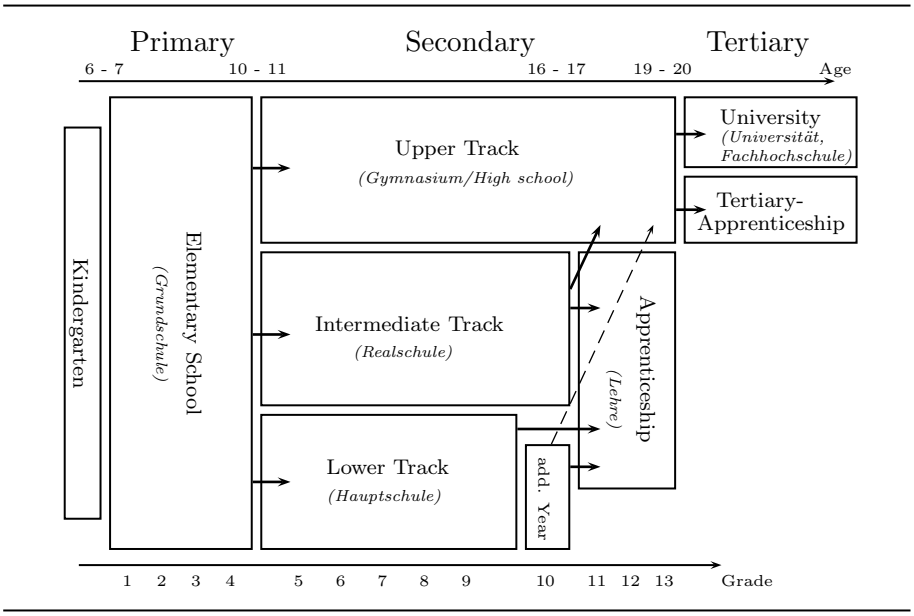
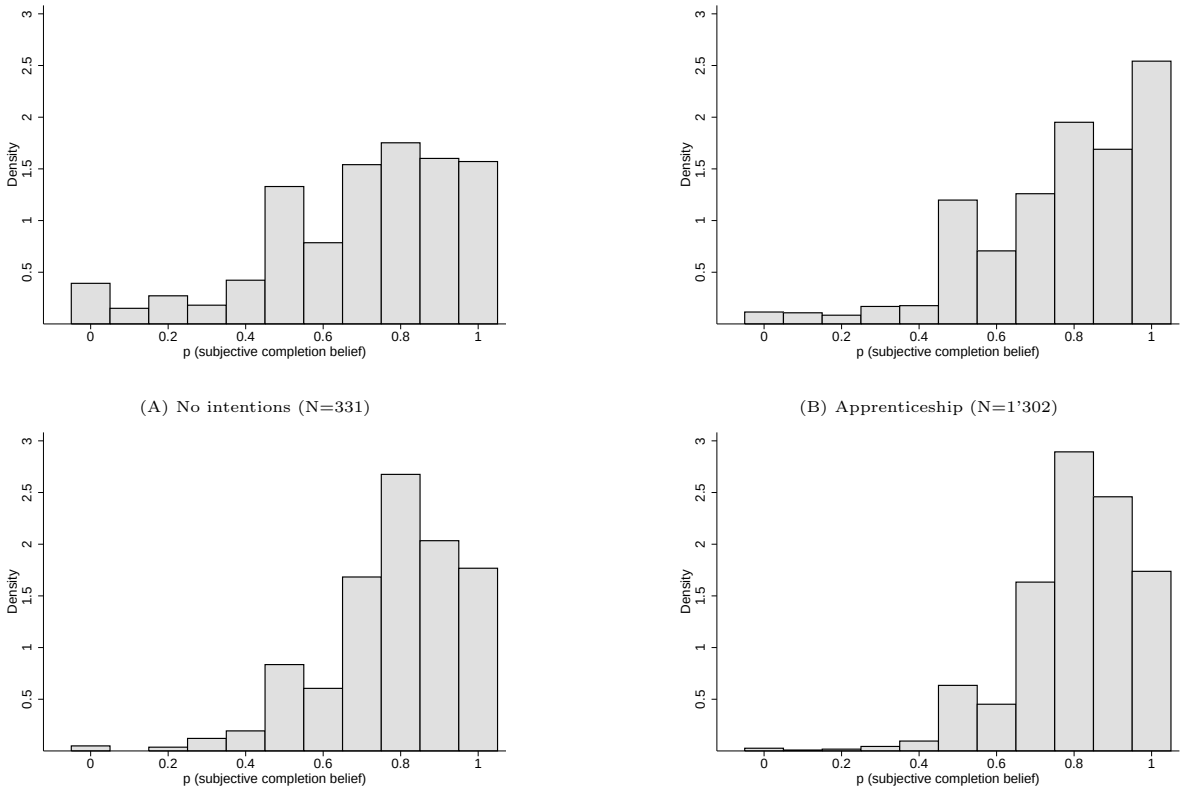


Figure 1: Simplification of the German education system

Source: Adaptation and extension of the overview provided by Wölfel and Heineck (2012).



(A) No intentions (N=331)

(B) Apprenticeship (N=1'302)

(C) Tertiary apprenticeship (N=826)

(D) University (N=1'151)

Figure 2: Subjective completion beliefs by intentions-to-invest
Source: SOEP 2000-2013, own calculations.

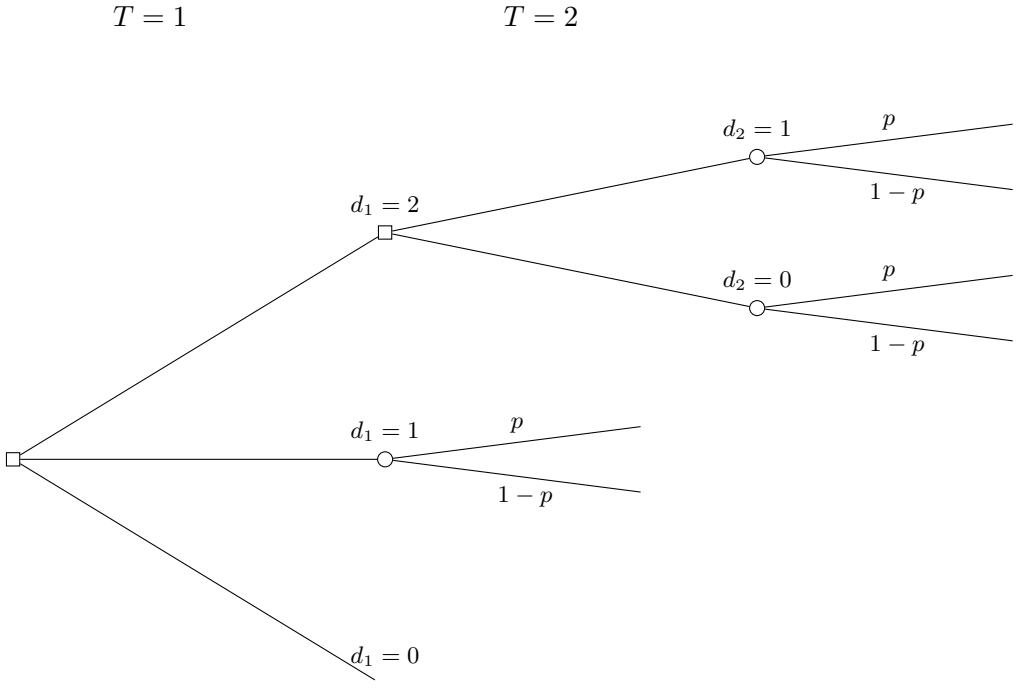


Figure 3: Sequential education decisions and timing of events

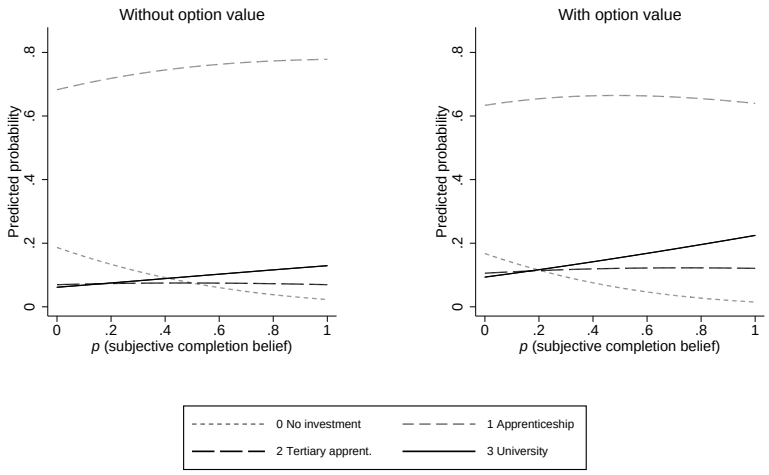


Figure 4: The role of the option value in a dynamic model of educational choice

Notes: Predicted probabilities constructed using estimates from Column (5) of Table 6 and evaluated at sample means.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

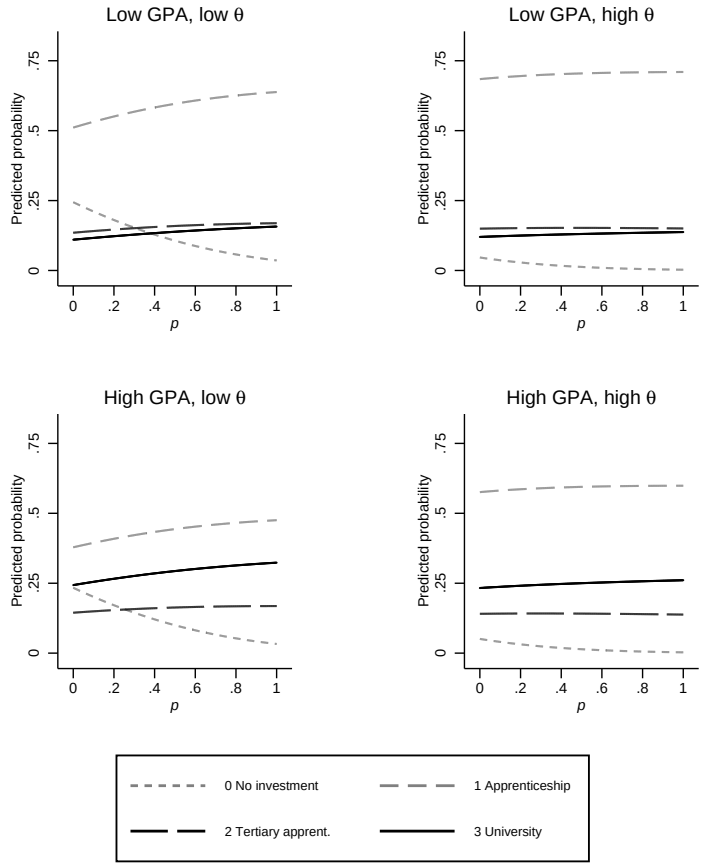


Figure 5: The role of academic ability and unobserved heterogeneity

Notes: Predicted probabilities constructed using estimates from Column (8) of Table 6, evaluated at sample means. High GPA = High $\theta = \Phi^{-1}(0.75)$, Low GPA = Low $\theta = \Phi^{-1}(0.25)$.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 1: Descriptive statistics by educational aspirations

	By aspiration level				Total
Variables	0	1	2	3	
p	0.692	0.768	0.782	0.805	0.776
	(0.258)	(0.220)	(0.176)	(0.155)	(0.198)
GPA (std)	-0.269	-0.288	-0.000	0.404	0.000
	(1.035)	(0.919)	(0.902)	(1.007)	(1.000)
Rec: Lower Track (yes/no)	0.166	0.265	0.044	0.034	0.132
	(0.373)	(0.441)	(0.204)	(0.181)	(0.338)
Rec: Intermediate Track (yes/no)	0.248	0.351	0.271	0.150	0.259
	(0.432)	(0.477)	(0.445)	(0.358)	(0.438)
Rec: High school (yes/no)	0.284	0.092	0.524	0.698	0.402
	(0.452)	(0.289)	(0.500)	(0.459)	(0.490)
In high school (yes/no)	0.272	0.015	0.596	0.767	0.411
	(0.446)	(0.123)	(0.491)	(0.423)	(0.492)
Locus of control (std)	-0.178	-0.169	0.082	0.195	0.003
	(1.014)	(0.963)	(0.861)	(0.862)	(0.928)
Risk attitudes (std)	-0.057	-0.006	-0.037	0.053	0.001
	(0.962)	(0.945)	(0.914)	(0.895)	(0.924)
Openness (std)	-0.178	-0.134	0.048	0.167	-0.000
	(0.990)	(0.956)	(0.879)	(0.926)	(0.942)
Agreeableness (std)	-0.103	-0.048	0.068	0.042	0.002
	(0.928)	(0.987)	(0.909)	(0.919)	(0.944)
Extraversion (std)	-0.135	-0.051	0.021	0.075	-0.002
	(0.948)	(0.914)	(0.919)	(0.989)	(0.945)
Neuroticism (std)	0.059	0.009	0.053	-0.066	-0.000
	(0.942)	(0.922)	(0.885)	(1.003)	(0.943)
Conscientiousness (std)	-0.132	0.053	-0.059	0.026	0.002
	(0.967)	(0.938)	(0.915)	(0.953)	(0.942)
Female (yes/no)	0.486	0.454	0.541	0.539	0.504
	(0.501)	(0.498)	(0.499)	(0.499)	(0.500)
Nr. of siblings	1.613	1.710	1.433	1.496	1.570
	(1.461)	(1.494)	(1.206)	(1.099)	(1.316)
Second-generation migrant (yes/no)	0.746	0.680	0.574	0.557	0.623
	(0.436)	(0.466)	(0.495)	(0.497)	(0.485)
Parent college-educated (yes/no)	0.199	0.101	0.306	0.495	0.283
	(0.400)	(0.301)	(0.461)	(0.500)	(0.450)
Parent cur. unemployed (yes/no)	0.124	0.160	0.087	0.045	0.103
	(0.330)	(0.367)	(0.282)	(0.208)	(0.304)
Log. net household income	10.019	9.890	10.624	10.855	10.377
	(2.231)	(2.216)	(1.358)	(1.295)	(1.834)
Cyclical youth unemployment (in Ror)	0.154	0.101	0.043	0.041	0.074
	(1.079)	(1.044)	(0.982)	(1.020)	(1.026)
Nr. of apprenticeship positions (in Ror)	98.380	98.544	98.538	99.368	98.791
	(4.906)	(5.261)	(5.600)	(5.124)	(5.279)
Nr. of students (in Ror)	23.700	22.711	24.156	25.730	24.095
	(14.204)	(14.354)	(13.991)	(14.091)	(14.223)
Nr. of high school graduates (in Ror)	26.081	25.755	27.289	27.758	26.775
	(6.313)	(6.526)	(6.218)	(7.064)	(6.673)
Nr. of Universities (in Ror)	10.789	9.620	10.916	11.381	10.585
	(10.304)	(9.666)	(10.037)	(9.988)	(9.938)
N	331 (9.17%)	1'302 (36.07%)	826 (22.88%)	1'151 (31.88%)	3'610

Note: Table presents sample means and standard deviations in brackets in total and by aspiration levels. Individual characteristics are assessed at the time of answering the Youth Questionnaire (with 17). GPA is the grade point average of German and Math grades, standardized and reversed, that higher values indicate better performance. Three indicators for school recommendations (with the age of 10), one indicator indicating if the student is currently in high school. Locus of control, openness, agreeableness, extraversion, neuroticism, and conscientiousness are principal components, std- stands for standardized to (0,1), where small deviations result from missings. We define second-generation migrants as having both parents born in a foreign country, parents college educated/unemployed if at least one has a college degree, is currently unemployed, cyclical component of youth unemployment in the region is extracted using the Hodrick-Prescott-Filter. The number of Universities in the region include all higher learning institutions.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 2: Determinants of subjective completion beliefs, OLS regressions

	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.029	0.029	0.029
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.025	0.026	0.025	0.028
	(0.014)	(0.014)	(0.014)	(0.015)
Rec: Intermediate Track (yes/no)	0.061	0.058	0.052	0.056
	(0.010)	(0.010)	(0.010)	(0.011)
Rec: High school (yes/no)	0.046	0.042	0.034	0.039
	(0.010)	(0.009)	(0.009)	(0.011)
In high school (yes/no)	0.005	0.003	-0.004	-0.005
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.022	0.019	0.019
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.008	0.005	0.005
		(0.004)	(0.004)	(0.004)
Openness (std)		0.005	0.004	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.008	0.008	0.008
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.016	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		0.001	0.002	0.001
		(0.004)	(0.004)	(0.004)
Conscientiousness (std)		0.031	0.034	0.034
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.013	-0.013
			(0.007)	(0.007)
Nr. siblings			-0.002	-0.003
			(0.002)	(0.002)
Second-generation migrant (yes/no)			-0.022	-0.010
			(0.007)	(0.013)
Parent college-educated (yes/no)			0.007	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			-0.002	0.001
			(0.013)	(0.013)
Log. net household income			0.007	0.007
			(0.002)	(0.002)
N	3,610	3,610	3,610	3,610
$\bar{p}$	0.776	0.776	0.776	0.776
$SD(p)$	0.198	0.198	0.198	0.198
adjR^2	0.052	0.107	0.115	0.116
Academic	+	+	+	+
F(pval)	30.470 (0.000)	20.679 (0.000)	17.564 (0.000)	17.029 (0.000)
Personality	-	+	+	+
F(pval)		24.168 (0.000)	24.626 (0.000)	24.328 (0.000)
Background	-	-	+	+
F(pval)			3.796 (0.000)	2.431 (0.013)
Labor market + FE	-	-	-	+
F(pval)				1.257 (0.184)

Note: Table presents coefficients, from linear regressions of subjective beliefs on varying sets of covariates, in (1) only on academic, (2) adds personality, (3) family background and individual characteristics, and (4) local labor market characteristics, region and time fixed effects (coefficients not presented). No recommendation is the base category, we include indicator variables for missing values in any of the covariates. Robust standard errors are given in brackets. We present the unconditional mean $\bar{p}$ and standard deviation $SD(p)$ of the dependent variable, the adjusted R^2, and joint significance tests. In the appendix we present analogous fractional response regressions.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 3: Effect of subjective completion beliefs on educational outcomes

	probit				probit eev			
					$\rho = .1$	$\rho = .3$	$\rho = .5$	$\rho = \hat{\rho}^o$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.921	0.809	0.726	0.716	0.612	0.383	0.120	0.561
	(0.142)	(0.146)	(0.148)	(0.150)	(0.149)	(0.141)	(0.129)	(0.168)
	[0.140]	[0.121]	[0.108]	[0.104]				
R^2_n	0.029	0.040	0.048	0.059				
$R^2_n(p)$	0.049	0.055	0.059	0.069				
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.776$, $SD(p) = 0.198$, $\hat{\rho}^o(se) = 0.147(0.003)$								
(B) Actual investment								
p	0.997	0.915	0.908	0.866	0.762	0.527	0.250	0.846
	(0.223)	(0.228)	(0.240)	(0.249)	(0.250)	(0.247)	(0.215)	(0.234)
	[0.069]	[0.062]	[0.056]	[0.044]				
R^2_n	0.087	0.100	0.122	0.184				
$R^2_n(p)$	0.113	0.121	0.141	0.199				
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.772$, $SD(p) = 0.201$, $\hat{\rho}^o(se) = 0.021(0.003)$								
(C) Actual investment, conditional on intentions								
p	0.901	0.845	0.836	0.726	0.622	0.393	0.129	0.716
	(0.256)	(0.262)	(0.272)	(0.276)	(0.277)	(0.269)	(0.240)	(0.141)
	[0.058]	[0.053]	[0.048]	[0.031]				
R^2_n	0.085	0.095	0.115	0.205				
$R^2_n(p)$	0.104	0.111	0.129	0.214				
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.781$, $SD(p) = 0.192$, $\hat{\rho}^o(se) = 0.010(0.002)$								
(D) Actual completion								
p	0.434	0.410	0.351	0.331	0.229	0.015	-0.214	0.326
	(0.181)	(0.185)	(0.189)	(0.192)	(0.190)	(0.190)	(0.171)	(0.094)
	[0.172]	[0.162]	[0.139]	[0.131]				
R^2_n	0.089	0.093	0.102	0.123				
$R^2_n(p)$	0.092	0.096	0.104	0.124				
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.769$, $SD(p) = 0.197$, $\hat{\rho}^o(se) = 0.005(0.003)$								
(E) Actual completion, conditional on intentions								
p	0.467	0.478	0.439	0.421	0.319	0.102	-0.135	0.418
	(0.198)	(0.202)	(0.206)	(0.210)	(0.210)	(0.191)	(0.197)	(0.083)
	[0.185]	[0.189]	[0.174]	[0.167]				
R^2_n	0.095	0.098	0.108	0.127				
$R^2_n(p)$	0.099	0.101	0.110	0.129				
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.778$, $SD(p) = 0.190$, $\hat{\rho}^o(se) = 0.004(0.003)$								
Academic	-	+	+	+	+	+	+	+
Personality	-	-	+	+	+	+	+	+
Family Background	-	-	-	+	+	+	+	+
Labor market	-	-	-	+	+	+	+	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. In the probit eev regressions we restrict the correlation between the errors to be 0.1, 0.3, 0.5 and to be equal to the selection-on-unobservables (the estimated is given by $\hat{\rho}^o$ along with its standard error). For each outcome in Panels (A) to (E), we present McFadden's pseudo-R^2 with and without p, and sample statistics for the varying subsamples. In the appendix we present analogous probit and bivariate probit regressions for dichotomized $p \geq 70\%$.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 4: Disaggregated intentions-to-invest, multinominal probit regressions

	(1)	(2)	(3)	(4)
Apprenticeship, $d = 1$				
p	1.045	1.136	1.038	1.068
	(0.206)	(0.214)	(0.218)	(0.221)
	[0.021(0.029)]	[0.081(0.029)]	[0.081(0.030)]	[0.089(0.030)]
Tertiary Apprenticeship, $d = 2$				
p	1.094	0.907	0.802	0.776
	(0.210)	(0.221)	(0.226)	(0.230)
	[-0.011(0.034)]	[-0.003(0.035)]	[0.002(0.036)]	[0.001(0.036)]
University, $d = 3$				
p	1.557	1.073	0.881	0.827
	(0.217)	(0.227)	(0.233)	(0.237)
	[0.141(0.035)]	[0.052(0.035)]	[0.030(0.036)]	[0.021(0.035)]
N	3,610	3,610	3,610	3,610
LR(pval)	54.900(0.000)	36.009(0.000)	26.476(0.000)	26.237(0.000)
Academic	-	+	+	+
Personality	-	-	+	+
Family Background	-	-	-	+
Labor market	-	-	-	+

Note: Table presents, multinominal probit regressions of the educational intention-to-invest: drop out, apprenticeship, tertiary apprenticeship, and university on subjective beliefs and varying sets of covariates in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. Robust standard errors in round, average marginal effect along (with standard errors) in squared (round) brackets. The Likelihood Ratio (LR)-statistic measures the significance of p across equations.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 5: Disaggregated actual investment and completion, conditional on intentions, probit regressions

	Actual investment				Actual completion			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Apprenticeship ($d_1 = 1$)								
p	0.902	1.055	1.007	1.043	0.794	0.737	0.645	0.559
	(0.256)	(0.266)	(0.269)	(0.275)	(0.271)	(0.273)	(0.276)	(0.278)
	[0.133]	[0.151]	[0.136]	[0.090]	[0.292]	[0.271]	[0.237]	[0.204]
R_n^2	0.033	0.044	0.073	0.175	0.051	0.060	0.069	0.106
$R_n^2(p)$	0.054	0.071	0.096	0.196	0.064	0.071	0.077	0.111
Sample:	$N = 883$, $\bar{d}_1 = 0.948$, $\bar{p} = 0.769$, $SD(p) = 0.218$				$N = 502$, $\bar{d}_1 = 0.669$, $\bar{p} = 0.760$, $SD(p) = 0.220$			
(B) Tertiary apprenticeship ($d_2 = 1$)								
p	0.362	0.455	0.673	0.784	0.409	0.358	0.374	0.669
	(0.368)	(0.379)	(0.412)	(0.416)	(0.549)	(0.564)	(0.578)	(0.619)
	[0.120]	[0.150]	[0.219]	[0.250]	[0.098]	[0.089]	[0.094]	[0.129]
R_n^2	0.079	0.084	0.107	0.137	0.140	0.139	0.148	0.223
$R_n^2(p)$	0.081	0.087	0.112	0.143	0.142	0.141	0.150	0.226
Sample:	$N = 456$, $\bar{d}_2 = 0.965$, $\bar{p} = 0.781$, $SD(p) = 0.179$				$N = 314$, $\bar{d}_2 = 0.557$, $\bar{p} = 0.779$, $SD(p) = 0.177$			
(C) University ($d_3 = 1$)								
p	0.789	0.431	0.308	0.038	0.980	0.665	0.894	0.656
	(0.367)	(0.388)	(0.405)	(0.409)	(0.487)	(0.512)	(0.538)	(0.546)
	[0.282]	[0.153]	[0.108]	[0.013]	[0.253]	[0.172]	[0.226]	[0.155]
R_n^2	0.140	0.186	0.219	0.275	0.177	0.194	0.223	0.252
$R_n^2(p)$	0.146	0.188	0.220	0.275	0.186	0.198	0.229	0.255
Sample:	$N = 580$, $\bar{d}_3 = 0.978$, $\bar{p} = 0.801$, $SD(p) = 0.154$				$N = 428$, $\bar{d}_3 = 0.397$, $\bar{p} = 0.799$, $SD(p) = 0.156$			
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit regressions of investment (1)-(4) and completion (5)-(8) on subjective completion beliefs and varying sets of covariate, in (1/5) on in high school, region and time fixed effects, (2/6) adds academic, (3/7) adds personality, (4/8) family background, individual, and local labor market characteristics. We present McFadden's pseudo-R^2 and sample statistics for the varying subsamples. For some regressions the numbers of observations are slightly reduced.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 6: Dynamic models of actual investment

	Dynamic model				Dyn. model with unobs. heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Apprenticeship ($d_1 = 1$)								
p	1.173	1.213	1.187	1.198	1.406	1.432	1.407	1.420
	(0.303)	(0.314)	(0.327)	(0.328)	(0.313)	(0.324)	(0.338)	(0.339)
θ					1.000	1.000	1.000	1.000
					(.)	(.)	(.)	(.)
High school ($d_1 = 2$)								
p	1.307	1.143	1.147	1.186	1.549	1.370	1.370	1.409
	(0.388)	(0.406)	(0.423)	(0.391)	(0.396)	(0.414)	(0.432)	(0.401)
θ					0.846	0.846	0.851	0.848
					(0.061)	(0.062)	(0.062)	(0.054)
University ($d_1 = 2, d_2 = 1$)								
p	0.462	0.105	0.063	0.084	0.462	0.103	0.062	0.082
	(0.259)	(0.273)	(0.282)	(0.301)	(0.260)	(0.273)	(0.282)	(0.300)
θ					-0.006	-0.006	-0.006	-0.007
					(0.047)	(0.047)	(0.047)	(0.049)
$\ln(\sigma_3)$				-0.719				-0.688
				(0.717)				(0.708)
N	2,116	2,116	2,116	2,116	2,116	2,116	2,116	2,116
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family+Labor market	-	-	-	+	-	-	-	+

Note: Table presents estimates of the model in equation (8). In the panel "Dynamic model", $\theta_i = 0$ for all i. The model in the panel "Dyn. model with unobs. heterogeneity" estimated by MSL with 100 random draws from $N(0, 1)$. The sets of covariates correspond to those in Table 3. Standard errors in parentheses. All regressions include an indicator for being in high school with 17, region and time fixed effects.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Supplementary Material

A Data

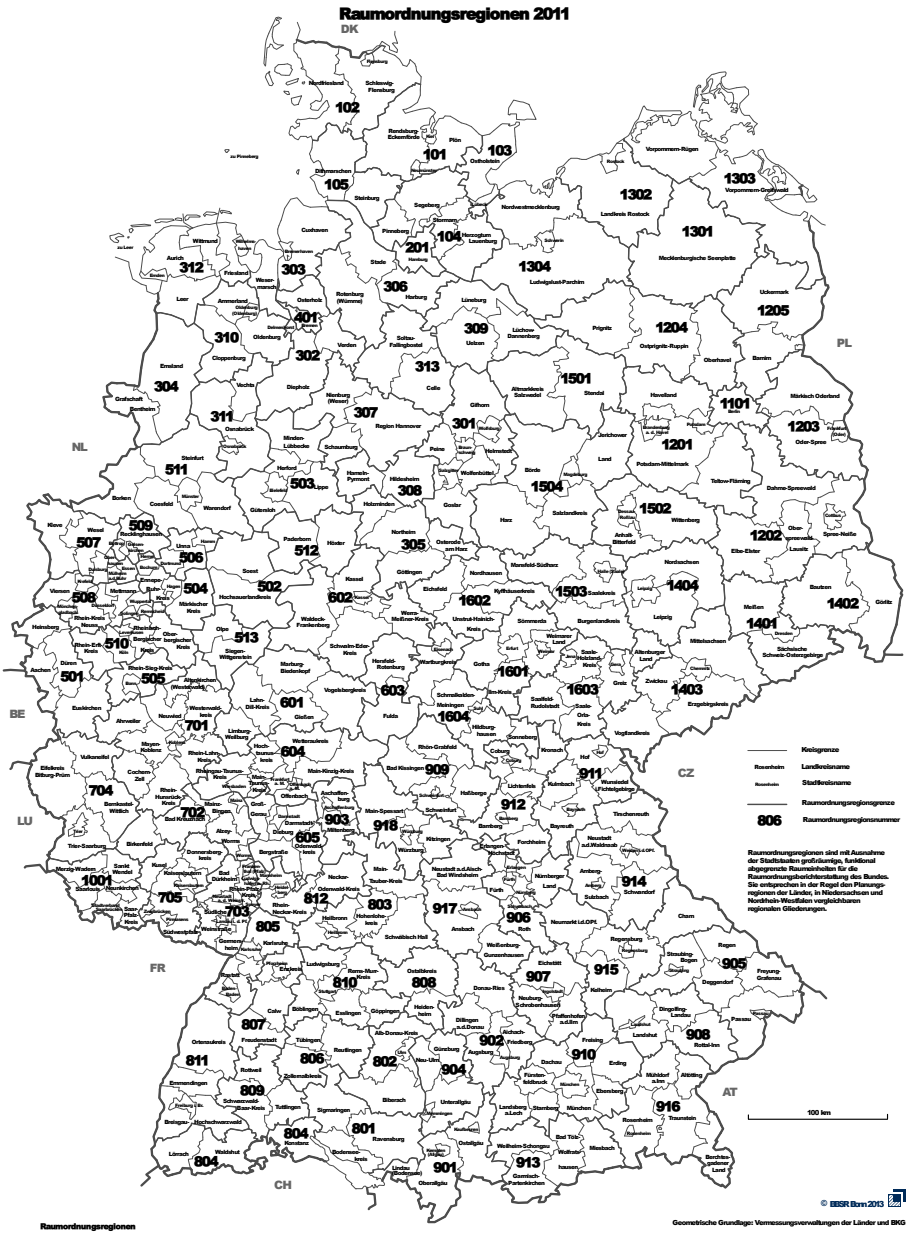


Figure A1: Local labor markets, 96 Raumordnungregionen [Ror]
Source: BBSR (2013)

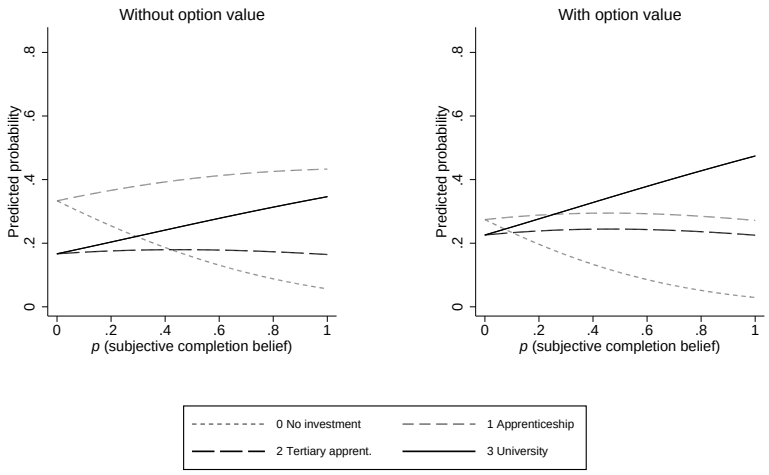


Figure A2: Robustness: The role of the option value in a dynamic model of educational choice

Notes: Predicted probabilities constructed using estimates from Column (5) of Table 6 and evaluated at $x'\beta_j = 0$, $j = 1, 2, 3$. *Source*: SOEP 2000-2013, INKAR 2012, own calculations.

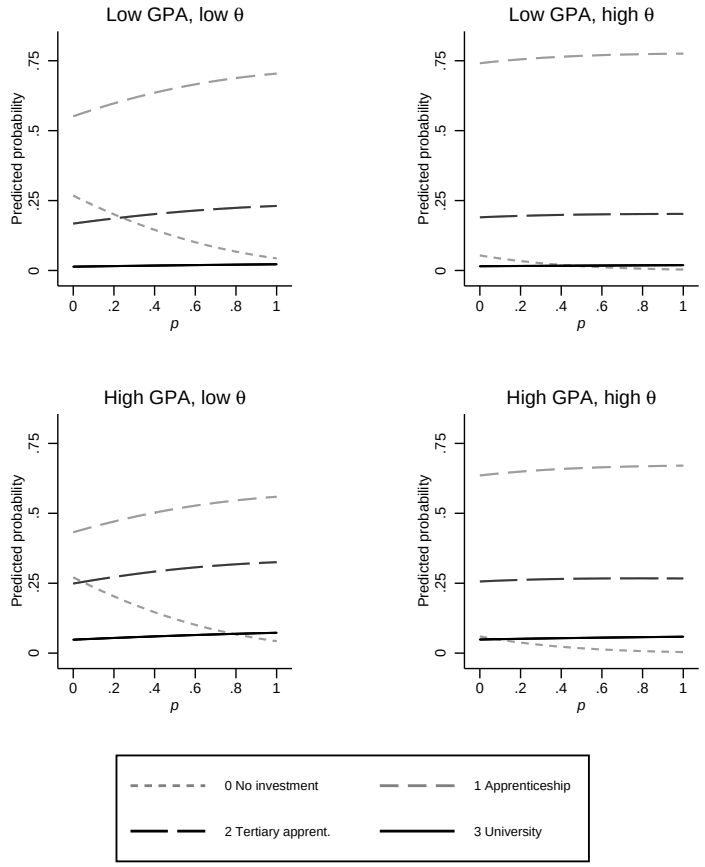


Figure A3: Robustness: The role of academic ability and unobserved heterogeneity

Notes: Predicted probabilities constructed using estimates from Column (8) of Table 6, evaluated at $x'\beta_j = 0$, $j = 1, 2, 3$. High GPA = High $\theta = \Phi^{-1}(0.75)$, Low GPA = Low $\theta = \Phi^{-1}(0.25)$.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A1: Descriptive statistics by subsample

	(A)	(B)	(C)	(D)	(F)
p	0.776	0.772	0.781	0.769	0.778
	(0.198)	(0.201)	(0.192)	(0.197)	(0.190)
GPA (std)	0.000	-0.018	0.012	0.041	0.077
	(1.000)	(1.019)	(1.009)	(1.024)	(1.016)
Rec: Lower Track (yes/no)	0.132	0.135	0.134	0.109	0.106
	(0.338)	(0.342)	(0.341)	(0.311)	(0.308)
Rec: Intermediate Track (yes/no)	0.259	0.272	0.270	0.247	0.249
	(0.438)	(0.445)	(0.444)	(0.431)	(0.433)
Rec: High school (yes/no)	0.402	0.347	0.360	0.384	0.400
	(0.490)	(0.476)	(0.480)	(0.487)	(0.490)
In high school (yes/no)	0.411	0.359	0.371	0.415	0.432
	(0.492)	(0.480)	(0.483)	(0.493)	(0.496)
Locus of control (std)	0.003	-0.038	-0.020	-0.053	-0.036
	(0.928)	(0.973)	(0.964)	(1.015)	(1.001)
Risk attitudes (std)	0.001	-0.085	-0.082	-0.153	-0.149
	(0.924)	(0.972)	(0.968)	(0.969)	(0.970)
Openness (std)	-0.000	-0.014	0.010	0.001	0.027
	(0.942)	(0.977)	(0.970)	(1.014)	(1.014)
Agreeableness (std)	0.002	-0.006	0.006	-0.011	0.006
	(0.944)	(0.984)	(0.980)	(0.995)	(1.002)
Extraversion (std)	-0.002	-0.010	0.009	-0.002	0.023
	(0.945)	(0.953)	(0.949)	(0.963)	(0.964)
Neuroticism (std)	-0.000	-0.015	-0.029	-0.027	-0.046
	(0.943)	(0.979)	(0.974)	(1.019)	(1.021)
Conscientiousness (std)	0.002	0.111	0.131	0.176	0.194
	(0.942)	(0.951)	(0.944)	(0.961)	(0.954)
Female (yes/no)	0.504	0.506	0.508	0.496	0.499
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Nr. siblings	1.570	1.639	1.630	1.633	1.622
	(1.316)	(1.339)	(1.323)	(1.335)	(1.322)
Second-generation migrant (yes/no)	0.623	0.739	0.731	0.843	0.840
	(0.485)	(0.439)	(0.443)	(0.364)	(0.367)
Parent college-educated (yes/no)	0.283	0.259	0.267	0.292	0.304
	(0.450)	(0.438)	(0.443)	(0.455)	(0.460)
Parent cur. unemployed (yes/no)	0.103	0.124	0.120	0.130	0.129
	(0.304)	(0.329)	(0.325)	(0.336)	(0.335)
Log. net household income	10.377	10.341	10.368	10.374	10.391
	(1.834)	(1.800)	(1.773)	(1.822)	(1.804)
Cyclical youth unemployment (in Ror)	0.074	0.211	0.211	0.267	0.260
	(1.026)	(1.034)	(1.035)	(1.083)	(1.082)
Nr. of apprenticeship positions (in Ror)	98.791	97.920	97.886	97.447	97.362
	(5.279)	(4.988)	(5.039)	(5.034)	(5.067)
Nr. of students (in Ror)	24.095	22.952	23.018	22.395	22.572
	(14.223)	(13.594)	(13.511)	(13.492)	(13.517)
Nr. of high school graduates (in Ror)	26.775	25.137	25.188	24.491	24.582
	(6.673)	(5.514)	(5.520)	(5.155)	(5.158)
Nr. of Universities (in Ror)	10.585	10.333	10.321	10.208	10.225
	(9.938)	(9.850)	(9.849)	(9.651)	(9.637)
N	3,610	2,116	1,919	1,372	1,255

Note: Table presents sample means and standard deviations in brackets in total and by subsample considered in Table 3.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A2: Determinants of subjective completion beliefs, fractional response regressions

	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.030	0.029	0.030
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.023	0.024	0.024	0.025
	(0.012)	(0.012)	(0.012)	(0.013)
Rec: Intermediate Track (yes/no)	0.057	0.054	0.049	0.051
	(0.009)	(0.009)	(0.009)	(0.010)
Rec: High school (yes/no)	0.045	0.040	0.033	0.036
	(0.009)	(0.009)	(0.009)	(0.010)
In high school (yes/no)	0.004	0.002	-0.005	-0.006
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.022	0.020	0.019
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.007	0.005	0.005
		(0.004)	(0.004)	(0.004)
Openness (std)		0.004	0.004	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.008	0.008	0.008
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.015	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		0.000	0.002	0.000
		(0.004)	(0.004)	(0.004)
Conscientiousness (std)		0.030	0.033	0.033
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.011	-0.012
			(0.007)	(0.007)
Nr. siblings			-0.002	-0.003
			(0.002)	(0.002)
Second-generation migrant (yes/no)			-0.023	-0.009
			(0.007)	(0.013)
Parent college-educated (yes/no)			0.007	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			-0.002	0.002
			(0.012)	(0.012)
Log. net household income			0.006	0.006
			(0.002)	(0.002)
N	3'610	3'610	3'610	3'610
$\bar{p}$	0.776	0.776	0.776	0.776
$SD(p)$	0.198	0.198	0.198	0.198
R_n^2	0.054	0.114	0.122	0.129
Academic	+	+	+	+
F(pval)	193.883 (0.000)	130.487 (0.000)	109.471 (0.000)	107.115 (0.000)
Personality	-	+	+	+
F(pval)		194.414(0.000)	201.546(0.000)	199.798(0.000)
Family Background	-	-	+	+
F(pval)			31.445(0.000)	19.705(0.012)
Labor market + FE	-	-	-	+
F(pval)				29.572(0.162)

Note: The Table presents Bernoulli pseudo-maximum likelihood with probit conditional expectation function, as proposed by Papke and Wooldridge (1996, 2008). We report marginal effects and robust standard errors in round brackets, our goodness of fit measure is a nonlinear R-squared measure and is calculated as the squared correlation coefficient between the estimated conditional expectation and the observed subjective beliefs: $R_n^2 = corr(\hat{p}, p)^2$, where $\hat{p} = \Phi(x'\hat{\beta})$ due to the probit specification. The regressions of subjective beliefs are presented on varying sets of covariates, in (1) only on academic, (2) adds personality, (3) family background and individual characteristics, and (4) local labor market characteristics, region and time fixed effects (coefficients not presented). We present the unconditional mean $\bar{p}$ and standard deviation $SD(p)$ of the dependent variable.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A3: Robustness: dichotomizing subjective beliefs ($p \geq 0.70$)

	probit				bivariate probit			
					$\rho = .05$	$\rho = .1$	$\rho = .2$	$\rho = .3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.380	0.321	0.280	0.274	0.188	0.103	-0.067	-0.236
	(0.066)	(0.069)	(0.069)	(0.070)	(0.070)	(0.070)	(0.069)	(0.067)
	[0.068]	[0.055]	[0.047]	[0.045]				
R^2_n	0.027	0.040	0.048	0.059				
$R^2_n(p)$	0.041	0.049	0.055	0.065				
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.793$, $SD(p) = 0.406$								
(B) Actual investment								
p	0.490	0.453	0.439	0.429	0.343	0.257	0.087	-0.084
	(0.107)	(0.111)	(0.114)	(0.120)	(0.119)	(0.119)	(0.117)	(0.115)
	[0.045]	[0.039]	[0.035]	[0.028]				
R^2_n	0.086	0.100	0.122	0.184				
$R^2_n(p)$	0.110	0.120	0.140	0.198				
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.789$, $SD(p) = 0.408$								
(C) Actual investment, conditional on intentions								
p	0.471	0.453	0.443	0.437	0.351	0.264	0.092	-0.080
	(0.119)	(0.122)	(0.125)	(0.131)	(0.131)	(0.130)	(0.129)	(0.126)
	[0.040]	[0.037]	[0.033]	[0.025]				
R^2_n	0.083	0.095	0.115	0.205				
$R^2_n(p)$	0.105	0.114	0.132	0.220				
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.805$, $SD(p) = 0.396$								
(D) Actual completion								
p	0.231	0.225	0.209	0.196	0.109	0.022	-0.153	-0.328
	(0.089)	(0.091)	(0.093)	(0.094)	(0.094)	(0.094)	(0.093)	(0.092)
	[0.092]	[0.089]	[0.083]	[0.078]				
R^2_n	0.087	0.093	0.102	0.123				
$R^2_n(p)$	0.091	0.097	0.105	0.125				
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.794$, $SD(p) = 0.405$								
(E) Actual completion, conditional on intentions								
p	0.259	0.271	0.269	0.259	0.171	0.083	-0.094	-0.271
	(0.097)	(0.099)	(0.100)	(0.102)	(0.102)	(0.101)	(0.101)	(0.099)
	[0.103]	[0.108]	[0.107]	[0.103]				
R^2_n	0.092	0.098	0.108	0.127				
$R^2_n(p)$	0.097	0.103	0.112	0.131				
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.809$, $SD(p) = 0.393$								
Academic	-	+	+	+	+	+	+	+
Personality	-	-	+	+	+	+	+	+
Family Background	-	-	-	+	+	+	+	+
Labor market	-	-	-	+	+	+	+	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. In the bivariate probit regressions we restrict the correlation between the errors to be 0.05, 0.1, 0.2, 0.3. For each outcome in Panels (A) to (E), we present McFadden's pseudo-R^2 with and without p, and sample statistics for the varying subsamples. In the appendix we present analogous probit and bivariate probit regressions.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A4: Robustness: GPA standardized within high school and using federal states fixed effects

	GPA, by hs attendance				Federal states fe			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.921	0.810	0.727	0.717	0.928	0.817	0.733	0.722
	(0.142)	(0.146)	(0.148)	(0.150)	(0.142)	(0.146)	(0.148)	(0.150)
	[0.140]	[0.122]	[0.108]	[0.104]	[0.141]	[0.123]	[0.109]	[0.106]
R_n^2	0.029	0.040	0.048	0.059	0.027	0.038	0.046	0.058
$R_n^2(p)$	0.049	0.055	0.059	0.069	0.048	0.054	0.058	0.069
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.776$, $SD(p) = 0.198$								
(B) Actual investment								
p	0.997	0.918	0.911	0.870	1.033	0.959	0.947	0.888
	(0.223)	(0.228)	(0.241)	(0.250)	(0.224)	(0.229)	(0.242)	(0.252)
	[0.069]	[0.062]	[0.056]	[0.044]	[0.071]	[0.064]	[0.057]	[0.045]
R_n^2	0.087	0.100	0.122	0.183	0.089	0.102	0.124	0.180
$R_n^2(p)$	0.113	0.121	0.141	0.199	0.117	0.124	0.144	0.197
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.772$, $SD(p) = 0.201$								
(C) Actual investment, conditional on intentions								
p	0.901	0.849	0.840	0.730	0.924	0.877	0.862	0.729
	(0.256)	(0.262)	(0.272)	(0.276)	(0.255)	(0.263)	(0.273)	(0.279)
	[0.058]	[0.053]	[0.048]	[0.031]	[0.058]	[0.054]	[0.049]	[0.032]
R_n^2	0.085	0.095	0.114	0.205	0.089	0.098	0.118	0.204
$R_n^2(p)$	0.104	0.110	0.129	0.214	0.108	0.115	0.133	0.213
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.781$, $SD(p) = 0.192$								
(D) Actual completion								
p	0.434	0.410	0.351	0.331	0.408	0.378	0.333	0.312
	(0.181)	(0.184)	(0.189)	(0.192)	(0.180)	(0.184)	(0.188)	(0.191)
	[0.172]	[0.162]	[0.139]	[0.131]	[0.162]	[0.150]	[0.132]	[0.124]
R_n^2	0.089	0.093	0.102	0.123	0.078	0.082	0.092	0.112
$R_n^2(p)$	0.092	0.096	0.104	0.124	0.081	0.085	0.093	0.113
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.769$, $SD(p) = 0.197$								
(E) Actual completion, conditional on intentions								
p	0.467	0.479	0.439	0.421	0.454	0.455	0.430	0.410
	(0.198)	(0.202)	(0.206)	(0.210)	(0.197)	(0.200)	(0.205)	(0.208)
	[0.185]	[0.189]	[0.174]	[0.167]	[0.180]	[0.180]	[0.170]	[0.162]
R_n^2	0.095	0.098	0.108	0.127	0.084	0.087	0.097	0.117
$R_n^2(p)$	0.099	0.101	0.110	0.129	0.088	0.090	0.099	0.119
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.778$, $SD(p) = 0.190$								
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A5: Robustness: GPA standardized within federal states and including a fifth order polynomial in GPA

	GPA, std by federal states				GPA, polynomial			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.921	0.811	0.727	0.718	0.921	0.807	0.724	0.716
	(0.142)	(0.146)	(0.148)	(0.150)	(0.142)	(0.146)	(0.149)	(0.150)
	[0.140]	[0.122]	[0.108]	[0.105]	[0.140]	[0.121]	[0.108]	[0.104]
R_n^2	0.029	0.040	0.048	0.058	0.029	0.041	0.049	0.060
$R_n^2(p)$	0.049	0.055	0.059	0.069	0.049	0.056	0.060	0.071
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.776$, $SD(p) = 0.198$								
(B) Actual investment								
p	0.997	0.923	0.915	0.865	0.997	0.889	0.883	0.839
	(0.223)	(0.228)	(0.240)	(0.249)	(0.223)	(0.227)	(0.240)	(0.250)
	[0.069]	[0.062]	[0.056]	[0.044]	[0.069]	[0.057]	[0.051]	[0.040]
R_n^2	0.087	0.100	0.122	0.184	0.087	0.109	0.132	0.192
$R_n^2(p)$	0.113	0.121	0.141	0.199	0.113	0.128	0.149	0.206
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.772$, $SD(p) = 0.201$								
(C) Actual investment, conditional on intentions								
p	0.901	0.858	0.847	0.729	0.901	0.838	0.832	0.719
	(0.256)	(0.261)	(0.272)	(0.275)	(0.256)	(0.263)	(0.275)	(0.282)
	[0.058]	[0.054]	[0.049]	[0.031]	[0.058]	[0.050]	[0.045]	[0.030]
R_n^2	0.085	0.095	0.114	0.205	0.085	0.102	0.123	0.212
$R_n^2(p)$	0.104	0.110	0.129	0.214	0.104	0.117	0.137	0.221
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.781$, $SD(p) = 0.192$								
(D) Actual completion								
p	0.434	0.412	0.354	0.334	0.434	0.421	0.368	0.352
	(0.181)	(0.185)	(0.189)	(0.192)	(0.181)	(0.185)	(0.190)	(0.193)
	[0.172]	[0.163]	[0.140]	[0.133]	[0.172]	[0.167]	[0.146]	[0.140]
R_n^2	0.089	0.093	0.102	0.123	0.089	0.097	0.106	0.127
$R_n^2(p)$	0.092	0.096	0.104	0.124	0.092	0.099	0.108	0.129
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.769$, $SD(p) = 0.197$								
(E) Actual completion, conditional on intentions								
p	0.467	0.479	0.440	0.423	0.467	0.498	0.463	0.445
	(0.198)	(0.202)	(0.206)	(0.210)	(0.198)	(0.203)	(0.207)	(0.211)
	[0.185]	[0.190]	[0.174]	[0.167]	[0.185]	[0.197]	[0.183]	[0.176]
R_n^2	0.095	0.098	0.108	0.127	0.095	0.102	0.111	0.131
$R_n^2(p)$	0.099	0.101	0.110	0.129	0.099	0.105	0.114	0.133
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.778$, $SD(p) = 0.190$								
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A6: Robustness: separate regressions by high school attendance

	Not in high school				In high school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.988	0.880	0.802	0.801	0.655	0.446	0.424	0.466
	(0.160)	(0.164)	(0.167)	(0.168)	(0.303)	(0.311)	(0.324)	(0.335)
	[0.178]	[0.158]	[0.142]	[0.139]	[0.071]	[0.048]	[0.043]	[0.042]
R_n^2	0.018	0.030	0.042	0.052	0.030	0.042	0.059	0.088
$R_n^2(p)$	0.045	0.051	0.058	0.068	0.036	0.044	0.062	0.091
N	2'125	2'125	2'125	2'125	1'485	1'476	1'476	1'473
(B) Actual investment								
p	1.018	0.993	1.002	0.992	1.122	0.557	-0.013	-1.082
	(0.243)	(0.249)	(0.261)	(0.277)	(0.622)	(0.605)	(0.726)	(0.966)
	[0.097]	[0.092]	[0.086]	[0.068]	[0.036]	[0.010]	[0.000]	[0.000]
R_n^2	0.067	0.080	0.103	0.183	0.078	0.187	0.328	0.464
$R_n^2(p)$	0.097	0.106	0.128	0.204	0.099	0.191	0.328	0.469
N	1356	1356	1356	1355	584	582	582	577
(C) Actual investment, conditional on intentions								
p	0.950	0.979	0.992	0.934	0.710	0.088	-0.634	-3.554
	(0.281)	(0.285)	(0.291)	(0.311)	(0.653)	(0.651)	(0.823)	(1.551)
	[0.080]	[0.080]	[0.074]	[0.044]	[0.024]	[0.002]	[-0.002]	[0.000]
R_n^2	0.081	0.090	0.115	0.249	0.071	0.183	0.330	0.505
$R_n^2(p)$	0.103	0.113	0.136	0.264	0.078	0.183	0.333	0.538
N	1207	1207	1207	1206	546	544	544	539
(D) Actual completion								
p	0.534	0.501	0.420	0.367	0.247	0.193	0.213	0.230
	(0.212)	(0.215)	(0.219)	(0.225)	(0.347)	(0.359)	(0.380)	(0.391)
	[0.203]	[0.190]	[0.159]	[0.139]	[0.097]	[0.076]	[0.084]	[0.088]
R_n^2	0.055	0.062	0.071	0.103	0.115	0.120	0.141	0.170
$R_n^2(p)$	0.061	0.067	0.075	0.105	0.116	0.120	0.141	0.171
N	802	802	802	801	570	570	568	564
(E) Actual completion, conditional on intentions								
p	0.490	0.500	0.441	0.401	0.460	0.411	0.414	0.468
	(0.238)	(0.241)	(0.246)	(0.252)	(0.362)	(0.374)	(0.394)	(0.406)
	[0.183]	[0.187]	[0.165]	[0.150]	[0.180]	[0.161]	[0.162]	[0.179]
R_n^2	0.063	0.067	0.076	0.107	0.110	0.114	0.132	0.163
$R_n^2(p)$	0.068	0.072	0.080	0.109	0.112	0.115	0.134	0.165
N	709	709	709	708	535	535	533	529
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A7: Variable definitions

Variables	Description	Age	Missings
Core variables	*Missing values in the core variables are dropped from the estimation sample.*		
p	Subjective completion belief is a elicited measure, it ranges from 0 to 1, in 0.1 steps.	17	58
GPA (std)	Average of German and Math grades, standardized over the sample population, as a robustness check we additionally standardize within educational track (cf. Table A4).	17	59
Educational outcomes:	From the longitudinal information we assess whether the student has started/completed a respective educational track.	17-31	-
$d \in \{0, 1, 2, 3\}$	Aspiration/Start/Complete, disaggregated by the tracks: drop out, apprenticeship, tertiary apprenticeship (high school and apprenticeship), and university (includes all higher learning institutions).		
$d_1 \in \{0, 1, 2\}$	First stage in structural model: drop out, apprenticeship, and high school.		
$d_2 \in \{0, 1\}$	Second stage in structural model: tertiary apprenticeship and university.		
Start apprenticeship	Not used in the analysis, all individuals that stared before are dropped from the estimation sample.	17	487
Still in school	Used in aspiration regressions, but dropped in the investment/completion analysis.	17	1,073
Academic variables			
Recommendations:	To visit a secondary-school track teachers evaluate the students (age the age of 10), the base category is no recommendation, three indicators for Lowest Track (yes/no), Intermediate Track (yes/no), and High school (yes/no)	17	249
In high school (yes/no)	An indicator whether the student is currently in high school when answering the youth questionnaire.	17	105
Personality variables	*We standardize the personality variables to mean 0 and standard deviation 1.*		
Locus of control (std)	First principal component of 10 questions, of which two are reversed.	17	459
Risk attitudes (std)	Assessed by a single question, ranging from 1-10.	17	306
Openness (std)	First principal component of 3 questions.	17	381
Agreeableness (std)	First principal component of 3 questions, of which one is reversed.	17	375
Extraversion (std)	First principal component of 3 questions, of which one is reversed.	17	378
Neuroticism (std)	First principal component of 3 questions, of which one is reversed.	17	378
Conscientiousness (std)	First principal component of 3 questions, of which one is reversed.	17	381

Individual and family characteristics	*Parental information, based on parents' questionnaires, are merged with the children's information.*		
Female (yes/no)	An indicator whether the individual is female.	17	
Nr. of siblings	Count of the number of siblings.	17	179
Second-generation migrant (yes/no)	An indicator whether the individual's parents are born in a foreign country, if information is missing recoded as second-generation migrant.	17	2,029
Parent college-educated (yes/no)	An indicator whether the individual has at least one college educated parent.	17	13
Parent cur. unemployed (yes/no)	An indicator whether the individual has at least one currently unemployed parent.	17	152
Log. net household income	Log of household pre-governmental income imputed by SOEP (0 income is treated as missing)	17	65
Fixed effects			
Year	Year of answering youth questionnaire, which is roughly identical to year of birth	17	
Region	Five regions based on federal states which are the level of educational-jurisdiction, cf. footnote 23 and Table A4	17	109
Regional labor market information	*Information from INKAR 2012/Statistical agency, merged onto the students residence with 17 and lagged by one year. Some are twice assessed for the estimation of the structural model, based on residence with 17 to avoid endogeneity due to moving (there are no missings as the location is always known at 179.*		
Cyclical youth unemployment	Cyclical component of local youth unemployment, extracted by HP-filter.	16/18	
Nr. of apprenticeship positions	Number of apprenticeship positions by all potential apprentices times 100.	16/18	
Nr. of students	Number of students enrolled in higher learning institutions by all residents in the age group times 1000.	16/18	
Nr. of high school graduates	Number of students with a high school degree in the region over all school-leavers times 1000.	16/18	
Nr. of universities	Count of higher learning institutions in the Ror, due to minimal variation over time we keep it constant.	16	

Note: Table presents variable descriptions and missing values for the baseline sample. All available individuals add up to 4,192, which then reduce to 3,610. The remaining missings are conditional on the estimation sample. All variables besides core variables are included in the estimation along with missing value indicators. More information on the regional indicators can be found under `http://www.inkar.de`